INTEGRATIVE GENOMICS AND NETWORK BIOLOGY IN LIVESTOCK AND OTHER DOMESTIC ANIMALS

EDITED BY: David E. MacHugh and Robert J. Schaefer

PUBLISHED IN: Frontiers in Genetics and Frontiers in Veterinary Science

frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88963-999-1
DOI 10.3389/978-2-88963-999-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INTEGRATIVE GENOMICS AND NETWORK BIOLOGY IN LIVESTOCK AND OTHER DOMESTIC ANIMALS

Topic Editors:
David E. MacHugh, University College Dublin, Ireland
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Citation: MacHugh, D. E., Schaefer, R. J., eds. (2020). Integrative Genomics and Network Biology in Livestock and Other Domestic Animals. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-999-1

Table of Contents

07 ***Defining Key Genes Regulating Morphogenesis of Apocrine Sweat Gland in Sheepskin***
Shaomei Li, Xinting Zheng, Yangfan Nie, Wenshuo Chen, Zhiwei Liu, Yingfeng Tao, Xuewen Hu, Yong Hu, Haisheng Qiao, Quanqing Qi, Quanbang Pei, Danzhuoma Cai, Mei Yu and Chunyan Mou

23 ***Genetic Basis of Phenotypic Differences Between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-Specific Expression in Their F1 Hybrids***
Yanhong Cao, Han Xu, Ran Li, Shan Gao, Ningbo Chen, Jun Luo and Yu Jiang

32 ***Detection of Co-expressed Pathway Modules Associated With Mineral Concentration and Meat Quality in Nelore Cattle***
Wellison J. S. Diniz, Gianluca Mazzoni, Luiz L. Coutinho, Priyanka Banerjee, Ludwig Geistlinger, Aline S. M. Cesar, Francesca Bertolini, Juliana Afonso, Priscila S. N. de Oliveira, Polyana C. Tizioto, Haja N. Kadarmideen and Luciana C. A. Regitano

44 ***Systems Biology Reveals* NR2F6 *and* TGFB1 *as Key Regulators of Feed Efficiency in Beef Cattle***
Pâmela A. Alexandre, Marina Naval-Sanchez, Laercio R. Porto-Neto, José Bento S. Ferraz, Antonio Reverter and Heidge Fukumasu

60 ***Genetic Regulation of Liver Metabolites and Transcripts Linking to Biochemical-Clinical Parameters***
Siriluck Ponsuksili, Nares Trakooljul, Frieder Hadlich, Karen Methling, Michael Lalk, Eduard Murani and Klaus Wimmers

75 ***An Epigenome-Wide DNA Methylation Map of Testis in Pigs for Study of Complex Traits***
Xiao Wang and Haja N. Kadarmideen

88 ***Dynamic Transcriptome Analysis Reveals Potential Long Non-coding RNAs Governing Postnatal Pineal Development in Pig***
Yalan Yang, Rong Zhou, Wentong Li, Ying Liu, Yanmin Zhang, Hong Ao, Hua Li and Kui Li

99 ***Genetic Contribution to Variation in Blood Calcium, Phosphorus, and Alkaline Phosphatase Activity in Pigs***
Henry Reyer, Michael Oster, Dörte Wittenburg, Eduard Murani, Siriluck Ponsuksili and Klaus Wimmers

111 ***Transcriptomic Profiling of Duodenal Epithelium Reveals Temporally Dynamic Impacts of Direct Duodenal Starch-Infusion During Dry Period of Dairy Cattle***
Cong-Jun Li, Shudai Lin, María Jose Ranilla-García and Ransom L. Baldwin VI

122 ***Co-Expression Networks Reveal Potential Regulatory Roles of miRNAs in Fatty Acid Composition of Nelore Cattle***
Priscila S.N. de Oliveira, Luiz L. Coutinho, Aline S.M. Cesar, Wellison J. da Silva Diniz, Marcela M. de Souza, Bruno G. Andrade, James E. Koltes, Gerson B. Mourão, Adhemar Zerlotini, James M. Reecy and Luciana C.A.

136 ***Differential microRNA Expression in Porcine Endometrium Involved in Remodeling and Angiogenesis That Contributes to Embryonic Implantation***
Linjun Hong, Ruize Liu, Xiwu Qiao, Xingwang Wang, Shouqi Wang, Jiaqi Li, Zhenfang Wu and Hao Zhang

150 ***Differentially Expressed MiRNAs and tRNA Genes Affect Host Homeostasis During Highly Pathogenic Porcine Reproductive and Respiratory Syndrome Virus Infections in Young Pigs***
Damarius S. Fleming and Laura C. Miller

163 ***Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle***
Joana Lima, Marc D. Auffret, Robert D. Stewart, Richard J. Dewhurst, Carol-Anne Duthie, Timothy J. Snelling, Alan W. Walker, Tom C. Freeman, Mick Watson and Rainer Roehe

181 ***Comparative Transcriptomic and Proteomic Analyses Identify Key Genes Associated With Milk Fat Traits in Chinese Holstein Cows***
Chenghao Zhou, Dan Shen, Cong Li, Wentao Cai, Shuli Liu, Hongwei Yin, Shaolei Shi, Mingyue Cao and Shengli Zhang

195 ***Comprehensive Profiles of mRNAs and miRNAs Reveal Molecular Characteristics of Multiple Organ Physiologies and Development in Pigs***
Muya Chen, Yi Long Yao, Yalan Yang, Min Zhu, Yijie Tang, Siyuan Liu, Kui Li and Zhonglin Tang

210 ***Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep***
Mazdak Salavati, Stephen J. Bush, Sergio Palma-Vera, Mary E. B. McCulloch, David A. Hume and Emily L. Clark

226 ***Genetic Parameters for Yolk Cholesterol and Transcriptional Evidence Indicate a Role of Lipoprotein Lipase in the Cholesterol Metabolism of the Chinese Wenchang Chicken***
Xingyong Chen, Wenjun Zhu, Yeye Du, Xue Liu and Zhaoyu Geng

239 ***Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response***
Krishnamoorthy Srikanth, Himansu Kumar, Woncheoul Park, Mijeong Byun, Dajeong Lim, Steve Kemp, Marinus F. W. te Pas, Jun-Mo Kim and Jong-Eun Park

254 ***Corrigendum: Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response***
Krishnamoorthy Srikanth, Himansu Kumar, Woncheoul Park, Mijeong Byun, Dajeong Lim, Steve Kemp, Marinus F. W. te Pas, Jun-Mo Kim and Jong-Eun Park

255 ***Identification of lncRNAs by RNA Sequencing Analysis During* in Vivo *Pre-Implantation Developmental Transformation in the Goat***
Ying-hui Ling, Qi Zheng, Yun-sheng Li, Meng-hua Sui, Hao Wu, Yun-hai Zhang, Ming-xing Chu, Yue-hui Ma, Fu-gui Fang and Li-na Xu

267 *Gene Expression and Fatty Acid Profiling in* Longissimus thoracis *Muscle, Subcutaneous Fat, and Liver of Light Lambs in Response to Concentrate or Alfalfa Grazing*
Elda Dervishi, Laura González-Calvo, Mireia Blanco, Margalida Joy, Pilar Sarto, R. Martin-Hernandez, Jose M. Ordovás, Magdalena Serrano and Jorge H. Calvo

283 *A Mini-Atlas of Gene Expression for the Domestic Goat (*Capra hircus*)*
Charity Muriuki, Stephen J. Bush, Mazdak Salavati, Mary E.B. McCulloch, Zofia M. Lisowski, Morris Agaba, Appolinaire Djikeng, David A. Hume and Emily L. Clark

299 *Analysis of the Progeny of Sibling Matings Reveals Regulatory Variation Impacting the Transcriptome of Immune Cells in Commercial Chickens*
Lucy Freem, Kim M. Summers, Almas A. Gheyas, Androniki Psifidi, Kay Boulton, Amanda MacCallum, Rakhi Harne, Jenny O'Dell, Stephen J. Bush and David A. Hume

319 *Conjoint Analysis of SMRT- and Illumina-Based RNA-Sequencing Data of* Fenneropenaeus chinensis *Provides Insight Into Sex-Biased Expression Genes Involved in Sexual Dimorphism*
Qiong Wang, Yuying He and Jian Li

330 *Genome-Wide Differential Expression Profiling of Ovarian circRNAs Associated With Litter Size in Pigs*
Gaoxiao Xu, Huifang Zhang, Xiao Li, Jianhong Hu, Gongshe Yang and Shiduo Sun

338 *Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle*
Wietje Nolte, Rosemarie Weikard, Ronald M. Brunner, Elke Albrecht, Harald M. Hammon, Antonio Reverter and Christa Kühn

357 *Expression Quantitative Trait Loci in Equine Skeletal Muscle Reveals Heritable Variation in Metabolism and the Training Responsive Transcriptome*
Gabriella Farries, Kenneth Bryan, Charlotte L. McGivney, Paul A. McGettigan, Katie F. Gough, John A. Browne, David E. MacHugh, Lisa Michelle Katz and Emmeline W. Hill

373 *Integrated Analysis of Methylome and Transcriptome Changes Reveals the Underlying Regulatory Signatures Driving Curly Wool Transformation in Chinese Zhongwei Goats*
Ping Xiao, Tao Zhong, Zhanfa Liu, Yangyang Ding, Weijun Guan, Xiaohong He, Yabin Pu, Lin Jiang, Yuehui Ma and Qianjun Zhao

387 *Dynamic Transcriptomic Analysis of Breast Muscle Development From the Embryonic to Post-hatching Periods in Chickens*
Jie Liu, Qiuxia Lei, Fuwei Li, Yan Zhou, Jinbo Gao, Wei Liu, Haixia Han and Dingguo Cao

399 *Integrated Hypothalamic Transcriptome Profiling Reveals the Reproductive Roles of mRNAs and miRNAs in Sheep*
Zhuangbiao Zhang, Jishun Tang, Ran Di, Qiuyue Liu, Xiangyu Wang, Shangquan Gan, Xiaosheng Zhang, Jinlong Zhang, Mingxing Chu and Wenping Hu

413 ***Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs***

Brittney N. Keel, Warren M. Snelling, Amanda K. Lindholm-Perry, William T. Oliver, Larry A. Kuehn and Gary A. Rohrer

425 ***Alveolar Macrophage Chromatin is Modified to Orchestrate Host Response to* Mycobacterium bovis *Infection***

Thomas J. Hall, Douglas Vernimmen, John A. Browne, Michael P. Mullen, Stephen V. Gordon, David E. MacHugh and Alan M. O'Doherty

439 ***Emerging Roles of Heat-Induced circRNAs Related to Lactogenesis in Lactating Sows***

Jiajie Sun, Haojie Zhang, Baoyu Hu, Yueqin Xie, Dongyang Wang, Jinzhi Zhang, Ting Chen, Junyi Luo, Songbo Wang, Qinyan Jiang, Qianyun Xi, Zujing Chen and Yongliang Zhang

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 January 2019
doi: 10.3389/fgene.2018.00739

Defining Key Genes Regulating Morphogenesis of Apocrine Sweat Gland in Sheepskin

*Shaomei Li[1], Xinting Zheng[1], Yangfan Nie[1], Wenshuo Chen[1], Zhiwei Liu[1], Yingfeng Tao[1], Xuewen Hu[1], Yong Hu[2], Haisheng Qiao[2], Quanqing Qi[3], Quanbang Pei[3], Danzhuoma Cai[4], Mei Yu[1] and Chunyan Mou[1]**

[1] Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [2] Qinghai Academy of Animal Science and Veterinary Medicine, Xining, China, [3] Sanjiaocheng Sheep Breeding Farm, Haibei, China, [4] Animal Husbandry and Veterinary Station, Haixi, China

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota, Twin Cities, United States

Reviewed by:
Shaojun Liu,
Hunan Normal University, China
Jian Xu,
Chinese Academy of Fishery Sciences (CAFS), China

***Correspondence:**
Chunyan Mou
chunyanmou@mail.hzau.edu.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 13 October 2018
Accepted: 22 December 2018
Published: 30 January 2019

Citation:
Li S, Zheng X, Nie Y, Chen W, Liu Z, Tao Y, Hu X, Hu Y, Qiao H, Qi Q, Pei Q, Cai D, Yu M and Mou C (2019) Defining Key Genes Regulating Morphogenesis of Apocrine Sweat Gland in Sheepskin. Front. Genet. 9:739. doi: 10.3389/fgene.2018.00739

The apocrine sweat gland is a unique skin appendage in humans compared to mouse and chicken models. The absence of apocrine sweat glands in chicken and murine skin largely restrains further understanding of the complexity of human skin biology and skin diseases, like hircismus. Sheep may serve as an additional system for skin appendage investigation owing to the distributions and histological similarities between the apocrine sweat glands of sheep trunk skin and human armpit skin. To understand the molecular mechanisms underlying morphogenesis of apocrine sweat glands in sheepskin, transcriptome analyses were conducted to reveal 1631 differentially expressed genes that were mainly enriched in three functional groups (cellular component, molecular function and biological process), particularly in gland, epithelial, hair follicle and skin development. There were 7 Gene Ontology (GO) terms enriched in epithelial cell migration and morphogenesis of branching epithelium that were potentially correlated with the wool follicle peg elongation. An additional 5 GO terms were enriched in gland morphogenesis (20 genes), gland development (42 genes), salivary gland morphogenesis and development (8 genes), branching involved in salivary gland morphogenesis (6 genes) and mammary gland epithelial cell differentiation (4 genes). The enriched gland-related genes and two Kyoto Encyclopedia of Genes and Genomes pathway genes (WNT and TGF-β) were potentially involved in the induction of apocrine sweat glands. Genes named *BMPR1A, BMP7, SMAD4, TGFB3, WIF1*, and *WNT10B* were selected to validate transcript expression by qRT-PCR. Immunohistochemistry was performed to localize markers for hair follicle (SOX2), skin fibroblast (PDGFRB), stem cells (SOX9) and BMP signaling (SMAD5) in sheepskin. SOX2 and PDGFRB were absent in apocrine sweat glands. SOX9 and SMAD5 were both observed in precursor cells of apocrine sweat glands and later in gland ducts. These results combined with the upregulation of BMP signaling genes indicate that apocrine sweat glands were originated from outer root sheath of primary wool follicle

and positively regulated by BMP signaling. This report established the primary network regulating early development of apocrine sweat glands in sheepskin and will facilitate the further understanding of histology and pathology of apocrine sweat glands in human and companion animal skin.

Keywords: sweat gland, wool follicle, skin, morphogenesis, WNT, TGF-β, transcriptome

INTRODUCTION

Human skin is the largest organ that covers the body surface and balances heat and protects against assaults from the environment. It contains different subtypes of appendages including hair follicles, nails, sebaceous glands and sweat glands that display diverse histological structures and regional localizations in different body parts. The apocrine and eccrine sweat glands are two types distributed across the human body. Though both sweat glands have similar structures consisting of ductal and secretory portions, they do have different functions and locations (Sato et al., 1989). The eccrine sweat glands are generally found on hairless body regions, especially on the palms and soles (Sato et al., 1989), with slim ducts and small secretory portions that secrete water and electrolytes directly to the surface of the human body (Lobitz and Dobson, 1961; Yanagawa et al., 1986). In contrast, the apocrine sweat glands are connected to the upper part of hair follicles in hairy regions such as axilla and perineum (Sato et al., 1989), with short and thick ducts and large secretory coils that release viscous liquid (water, electrolytes, protein, lipids, and steroids to the opening of hair follicles) (Sato et al., 1989; Wilke et al., 2007). Under disease conditions with hircismus, the secretions from apocrine sweat glands are turned from originally odorless to odorous compounds by bacterial enzymes on the skin's surface (Shehadeh and Kligman, 1963; Preti and Leyden, 2010). A functional allele (538G > A) in the ATP-binding cassette C11 (ABCC11) gene was reported to highly associate with human earwax type (wet or dry) and axillary odor (Yoshiura et al., 2006; Toyoda et al., 2009, 2017; Martin et al., 2010). The increased expression of ABCC11 in apocrine sweat glands was detected more in the myoepithelial cells of the secretory portions in individuals with GG genotype than those of AA genotype (Toyoda et al., 2017). Though these reports suggested an interesting correlation between the ABCC11 gene and the axillary odor caused by apocrine sweat glands in human skin, the actual mechanisms underlying it remain unknown. More information related to apocrine sweat glands may assist the diagnosis and even practical treatment of this skin disease.

The apocrine sweat gland marks a big difference in the skin between humans and animal models (chicken and mouse) in that murine and chicken skin are exclusively lacking in apocrine sweat glands. Additionally, its absence in murine and chicken skin restrains its related investigations. Sheepskin may represent an additional system to gain basic information about apocrine sweat glands. The histological structure of apocrine sweat glands in the armpit skin are similar to those of sheep body skin (Sato et al., 1989; Rogers, 2006). Hence, the general knowledge of sheepskin would facilitate further understanding of human apocrine sweat glands under normal and diseased conditions.

Previous studies regarding sweat glands mainly focused on the eccrine sweat glands by detecting *KRT* gene expression in human embryos (Hashimoto et al., 1965; Sun et al., 1979; Moll and Moll, 1992) and elucidating the molecular mechanisms of morphogenesis and development in mouse models (Kunisada et al., 2009; Cui et al., 2014; Lu et al., 2016). Several signaling pathways including wingless-related integration site (WNT), ectodysplasin A receptor (EDAR), bone morphogenetic proteins (BMP), sonic hedgehog (SHH), were shown to regulate the initiation and maturation of eccrine sweat glands (Kunisada et al., 2009; Cui et al., 2014; Lu et al., 2016).

In *Eda*-null (tabby) mice, no eccrine sweat glands were formed throughout the embryonic stage in mouse paw skin (Kunisada et al., 2009). *β-catenin* conditional knockout mice showed complete blockage of eccrine sweat gland formation from E15.5 to birth before the unexpected death of the mice (Cui et al., 2014). *Wnt10a* mutant mice developed normal prenatal eccrine sweat gland germs but failed to form sweat ducts postnatally (Xu et al., 2017). Hence, *Wnt10a*/β-*catenin* mainly regulates the maturation of eccrine sweat glands in postnatal life. The BMP pathway has been reported to play a positive role in determining the glandular fate during the induction stage of eccrine sweat gland. In *Bmpr1a* conditional knockout mice, the eccrine sweat glands were converted to hair follicle-like structures (Lu et al., 2016) and the density of eccrine sweat glands was reduced in *Bmp5* null mouse skin (Lu et al., 2016). The cross-talk of BMP and SHH spatiotemporally determined the subtypes of skin appendages, either hair follicles or eccrine sweat glands. A high BMP signal in mesenchyme and a low SHH signal in the epidermis engaged the glandular fate decision just before the initiation of eccrine sweat gland development (Lu et al., 2016). This mechanism was also observed in other ectodermal glands (mammary and meibomian) and chicken digestive epithelia formation (Roberts et al., 1998; Narita et al., 2000; Mayer et al., 2008; Huang et al., 2009a). These findings highly suggest that inhibiting BMP signaling favors hair follicle cell fates, whereas active BMP signaling promotes glandular cell fates. In addition, the eccrine sweat gland density was also shown to be determined by the expression of homeodomain transcription factor engrailed 1 (*En1*) in murine footpad skin (Kamberov et al., 2015).

Though the eccrine sweat glands gained most of the research interest, the understanding of the apocrine sweat glands was relatively restricted to the physiological and pathological descriptions of humans or companion animals, owing to the absence of apocrine sweat glands in mouse and chicken skin (Leyden et al., 1981; Kalaher et al., 1990; Morandi et al., 2005; Baharak et al., 2012; Fujiwara-Igarashi et al., 2017). Until now,

there is little in the literature related to the morphogenesis and development of apocrine sweat glands.

Previously, sheepskin attracted researchers to decipher the regulatory mechanisms of wool follicle and skin development in embryonic stages and in postnatal seasonal wool growth. The induction of primary wool follicles in coarse wool sheepskin during early embryonic stages and the morphogenesis of secondary wool follicles in merino sheepskin were investigated by exploring the interaction network of long non-coding RNAs (lncRNAs) and mRNAs, including a series of lncRNAs and WNT, BMP, EDAR, and FGF signaling pathways (Yue et al., 2016; Nie et al., 2018). The microRNA profiles identified candidates (miRNA-143, miRNA-10a, let-7i) potentially regulating the different wool follicle growth patterns with small, medium or large waves in Hu sheepskin, and a series of new microRNAs during the wool follicle seasonal growth cycling in sheepskin (Liu et al., 2013, 2014; Lv et al., 2016; Gao et al., 2017). These studies focused mainly on the morphogenesis of primary and secondary wool follicles and the regulation of wool fiber thickness. The existence of apocrine sweat glands in sheepskin is a great advantage for obtaining a deep understanding of the complexity of skin biology. Our current study aimed to explore the dynamic gene regulatory network and potential candidate genes governing the apocrine sweat gland induction in skin using sheepskin as a model system. This result will add general knowledge regarding the histological and molecular changes during the apocrine sweat gland morphogenesis and contribute to the further understanding of apocrine sweat gland development in skin of normal or diseased human or companion animals.

MATERIALS AND METHODS

Experimental Animals

Coarse wool sheep (Tibetan carpet wool sheep) fetuses were randomly collected from a local abattoir in Qinghai Province of China as described previously (Nie et al., 2018). Briefly, the discarded fetuses were rescued and immediately placed in PBS. The dorsal skin was dissected and divided into two parts. One part was fixed in 4% paraformaldehyde at 4℃ and the other part was frozen in liquid nitrogen for RNA extraction. The individuals (approximately 120 individuals) at unspecified embryonic stages were randomly collected for the determination of developmental stage by H&E (hematoxylin and eosin) staining. All experiments on animals were approved by the Standing Committee of Hubei People's Congress and the ethics committee of Huazhong Agricultural University.

H&E Staining

To identify the developmental stages of wool follicles and apocrine sweat glands in embryonic sheepskin, a series of fixed sheep dorsal skin samples were dehydrated with gradient alcohol, processed in paraffin and cut into 5 μm sections, according to the standard procedures. Then dorsal skin sections were processed into dewaxing and H&E (hematoxylin and eosin) staining. The stained skin sections were photographed and grouped into different developmental stages based on the structures of wool follicles and apocrine sweat glands as described in previous reports (Rogers, 2006).

Transcriptome Sequencing and Differentially Expressed Genes Analyses

Total RNA was extracted using TRIzol reagent from six sheepskin samples and RNA integrity was assessed using the RNA Nano 6000 Assay Kit with the Bioanalyzer 2100 system (Agilent Technologies, CA, United States). The sequencing library was constructed at Novogene (Beijing, China) using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) following the manufacturer's procedures. The library quality was assessed on the Agilent Bioanalyzer 2100 system. After cluster generation, the libraries were sequenced on an Illumina Hiseq platform (Hiseq X ten) with 150 bp paired-end reads.

The original sequenced reads were evaluated for data quality and then clean reads were mapped to the sheep reference genome (version: Oarv3.1) by Hisat2. HTSeq v0.9.1 was used to count the reads numbers mapped to each gene. The FPKM of each gene was then calculated to estimate gene expression level (Trapnell et al., 2010). After standardizing and testing the read counts, the differentially expressed genes (DEGs) were obtained. For biological replicates, genes with an adjusted P-value < 0.05 and $|\log2 \text{(Fold change)}| > 1$ were set as the threshold for differential expression (**Supplementary Table S1**).

GO Term, KEGG Enrichment and PPI Analyses of Differentially Expressed Genes

Gene Ontology (GO) enrichment analyses of DEGs were implemented by the GOseq R package. GO terms with corrected P-values less than 0.05 were considered significantly enriched by differential expressed genes (**Supplementary Table S2**). KOBAS software was used to test the statistical enrichment of differential expression genes in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (**Supplementary Table S3**). Protein–protein interaction analyses of DEGs were based on the commonly used STRING database, then Cytoscape software was used to realize the visualization of the interaction network (Shannon et al., 2003).

Quantitative Real-Time PCR (qRT-PCR) Validation

Several differentially expressed mRNAs were selected and confirmed by qRT-PCR with GAPDH used as an internal reference. qRT-PCR was carried out with a Roche LightCyclerR 96 using iTaqTM Universal SYBRRGreen Supermix (Bio-Rad, United States). The amplification procedures were held at 95℃ for 5 min initially, followed by 45 cycles of 95℃ for 15 s and 60℃ for 1 min. Quantification of mRNAs was performed using the $2^{-\Delta\Delta Ct}$ method with average cycle thresholds. The qRT-PCR data were generated from three independent

samples per stage and statistically analyzed using Student's *t*-test ($n \geq 3$).

Immunohistochemistry

Immunohistochemistry was applied to detect the expression pattern of skin appendage markers. The skin was dehydrated with ethanol, embedded in paraffin and sectioned at 5 to 6 μm thickness. The sections were dewaxed, processed to antigen retrieval and incubated with primary antibodies (Sox2, mouse, Santa Cruz,1:200; Sox9, mouse, Abcam, 1:200; pSmad5, Rabbit, Abcam,1:800; Pdgfrb, Rabbit, Abcam, 1:400) at 4°C overnight. The secondary antibody from the immunological kit (Proteintech, China) was incubated for 1 h at room temperature. Visualization was performed by using DAB staining (1:50) followed by hematoxylin counter-staining. Experiments were repeated at least twice.

RESULTS

The Morphological Characterization of Developing Wool Follicles and Apocrine Sweat Glands in Coarse Wool Sheep Back Skin

In this study, the Tibetan carpet wool sheep, a typical coarse wool sheep, was chosen for detailed investigation of the early development of apocrine sweat glands in skin. A series of sheep back skin sections were used to determine the induction and morphogenesis of apocrine sweat glands. The wool follicles and apocrine sweat glands were observed to occur sequentially based on histological H&E stain (**Figures 1A–I**). From the homogeneous thin skin layers (**Figure 1A**) to the appearance of skin appendages (**Figure 1I**), the obvious morphological changes were the thickening of the dermal and epidermal layers and the occurrence of wool follicle pegs, apocrine sweat glands and sebaceous glands. The first appendage that appeared in skin was the primary wool follicle that developed the placode and associated dermal condensation (**Figure 1B**) from the thin skin (**Figure 1A**), later sequentially grew downward to the dermis, elongated and progressed to maturation (**Figures 1C–I**). The first sign of apocrine sweat glands was observed as several cells were tightly packed and located on lateral side of the epidermal compartment of the primary wool follicle peg, indicating the occurrence of the precursor/progenitor cells of apocrine sweat glands (**Figure 1C**). As the cells proliferated and differentiated, these small cell patches gradually developed a small germ and later extended into the dermis to form the slim and long gland duct (**Figures 1E–I**). The elongation of gland ducts was directed at the angles that were closely parallel to wool follicle pegs as detected in **Figures 1G–I**. At this stage, the dermal condensates of primary wool follicles were encapsulated to become dermal papilla, and the sweat gland duct cavity was visible and surrounded by two layers of cells (**Figure 1I**). These stained sections clearly stated the initiation, budding, elongation and ductal cavity formation of apocrine sweat glands in prenatal coarse wool sheepskin.

Differentially Expressed Genes Involved in the Elongation of Wool Follicle Peg and Induction of Apocrine Sweat Glands in Coarse Wool Sheepskin

To understand the induction of apocrine sweat glands, two particular stages that corresponded to the pre-gland (stage TF1b, **Figure 1B**) and gland budding stage (stage TF2a, **Figure 1E**) to form ductal portions (Rogers, 2006) were selected to perform RNA sequencing. The morphological changes between two selected stages were the epidermal and dermal thickening, follicle germ elongation and apocrine sweat gland budding.

The sequencing data were processed for bioinformatics analyses. The criteria set up to enrich DEGs was $|\log2 \text{ fold change}| > 1$ and $P < 0.05$. A total of 1631 genes including 774 upregulated and 857 downregulated genes exhibited significant expression changes in the stage TF2a vs. the stage TF1b target group (**Figure 2A**). All the DEGs were compared with the published gene lists enriched in different compartments of P5 mouse skin (Sennett et al., 2015). The overlapping genes between these two data sets were picked up and presented in **Table 1**. Since the mouse back skin contains no sweat glands, the listed genes shared between mouse and sheep back skin were shown to highly associate with skin and hair/wool follicle development represented by 22 genes for the epidermis, 8 genes for dermal fibroblasts, 12 genes for the outer root sheath, 3 genes for matrix, 19 genes for melanocyte, 34 genes for dermal papilla, 10 genes for transit amplifying cells and 12 genes for hair follicle stem cells (**Table 1**). These results suggest that the regulatory genes for prenatal wool follicles and sheepskin were partially conserved with those of postnatal murine hair follicles and skin.

Enriched GO Term and KEGG Pathway Analyses of the Differentially Expressed Genes

The 1631 DEGs were processed to GO term and KEGG enrichment analyses. The most enriched GO terms were biological process, cellular component and molecular function, including organelle (represented by CST6, APOA1, and MKI67), gene expression (represented by MST1) and structural molecule activity (represented by TUBA4A) (**Figure 2B**). Further details revealed that the GO terms were highly enriched in three categories: hair follicle and skin development, gland development and epithelial development. Two GO terms were highly associated with hair follicle development (14 genes enriched) and skin development (33 genes enriched) (**Table 2**). The enrichment of 5 GO terms for epithelial differentiation, migration and the branching process represented mammary gland epithelial cell differentiation (4 genes enriched), epithelial cell migration (27 genes enriched), morphogenesis of a branching epithelium (24 genes enriched), morphogenesis of an epithelial fold (8 genes enriched), morphogenesis of an epithelial bud (5 genes enriched), embryonic epithelial tube formation (23 genes enriched) and branching morphogenesis of an epithelial tube (21 genes enriched) (**Table 2**). Most of the enriched genes showed an upregulated expression trend. These terms highly suggest

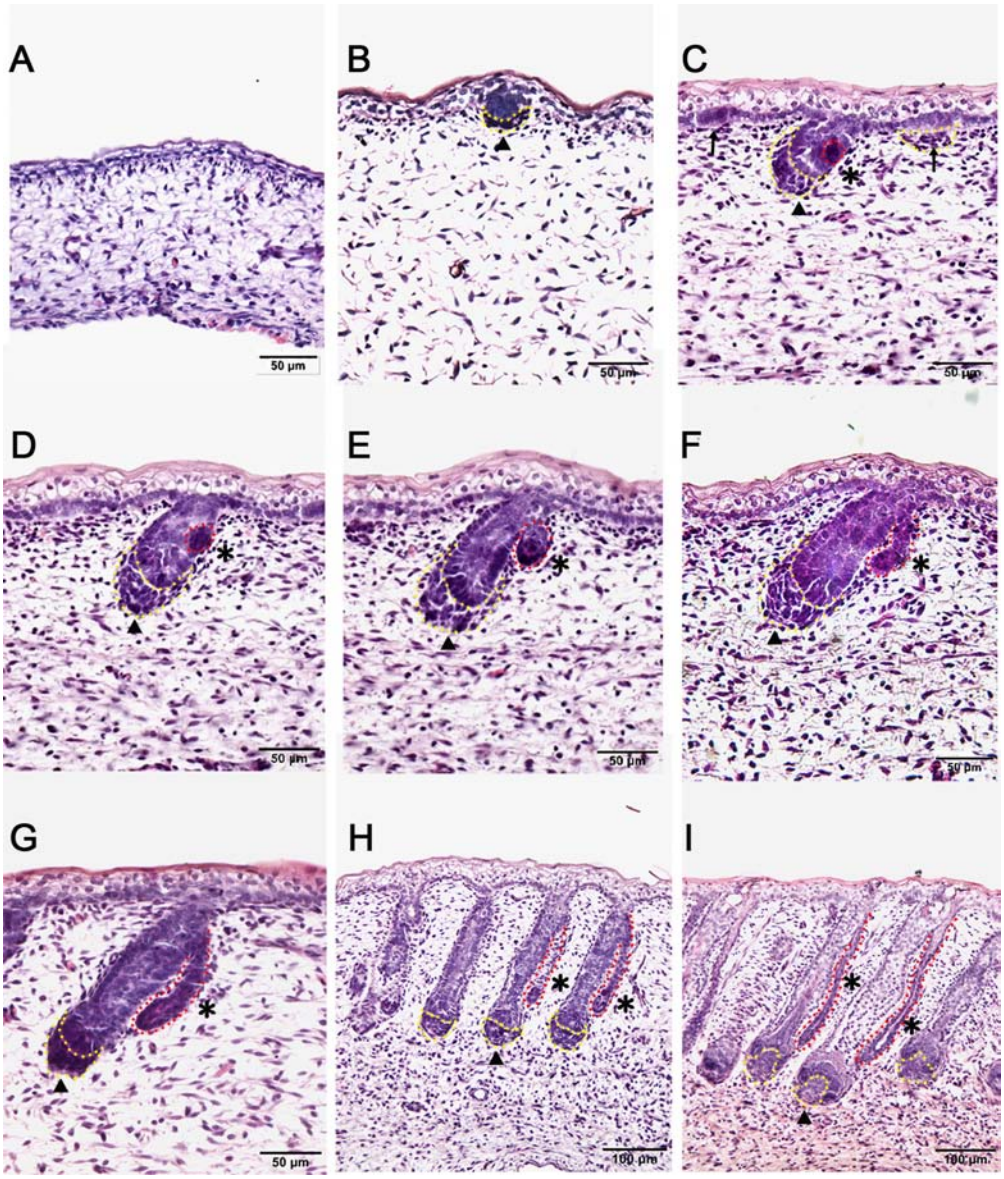


FIGURE 1 | Dynamic diagram of morphogenesis and development of wool follicles and apocrine sweat glands in embryonic coarse wool sheep. **(A)** The skin is homogeneous with thin epidermal and dermal layers before the initiation of skin appendages; **(B)** The first skin appendage, the primary wool follicle, is initiated to form epidermal placodes and associated dermal condensates as marked in a yellow dashed line; **(C)** The second skin appendage, the apocrine sweat gland is inducted at the lateral side of primary wool follicle epidermal peg. At the same time, the secondary wool follicle is initiated in between the primary wool follicles. The precursor/progenitor cells of apocrine sweat glands are marked with a red dashed line; **(D)** The precursor/progenitor cells of apocrine sweat glands exhibit early signs of the branching point from the lateral side of the outer root sheath of primary wool follicles; **(E)** The germ or small ductal bud of apocrine sweat glands protrudes and forms from the upper part of primary wool follicles. The primary and secondary wool follicles continue to grow downward into the dermis; **(F–H)** The apocrine sweat gland and wool follicle gradually elongate and extend into the dermis. The apocrine sweat gland grows and extends from germ to be a slim duct-like structure at the angle parallel to the primary wool follicle peg; **(I)** The primary wool follicle becomes mature with clear dermal papilla and matrix compared to the previous stage (Figure H). The ductal portion of the apocrine sweat gland gradually extends and forms a slim tube-like structure with an emerging cavity as indicated as asterisk. **A–G** Bar, 50 μm; **H** and **I** Bar, 100 μm.*, apocrine sweat gland; ▲, dermal condensates or dermal papilla; ↑, secondary hair follicle; dashed line in yellow, dermal papilla or dermal condensate; dashed line in red, apocrine sweat gland.

that during the two stages applied for RNA sequencing, the epithelia displayed the prominent biological functions, either for the elongation of the epidermal compartments of wool follicle pegs or the morphogenesis of the apocrine sweat gland ducts. Interestingly, a total of 5 GO terms were shown to regulate the gland morphogenesis (20 genes enriched), gland development (42 genes enriched), salivary gland morphogenesis and development (8 genes enriched), branching involved in salivary gland morphogenesis (6 genes enriched) and mammary gland epithelial cell differentiation (4 genes enriched).

TABLE 1 | The overlapped differentially expressed genes (DEGs) specific for different compartments of skin are presented by comparing the dataset of sheep prenatal skin with that of murine P5 dorsal skin.

Compartments of skin	Number	Genes
Epi	22	ANXA4 CST6 DEFB1 EMC9 IL20RA KLF3 KRTDAP LGALS3 LRP4 LY6D MGST2 MST1 NET1 PARP10 S100A11 SBSN STAP2 TGFBI TSPO TST TUBA4A WNT16
ORS	12	DAPK1 FLRT2 FNDC1 GPC4 LAMA5 LGR5 LHX2 MPDZ MYH9 PLA2G7 SOX9 TIMP3
Mx	3	FABP5 TGM1 RASGEF1B
Mc	19	ARSA B2M COMMD4 DKK3 DTNBP1 ETS1 GLUL LBH LITAF NKD1 NPC2 PRDX3 RAB5B SEMA6A STX7 TUBB3 VAT1 VPS26A ZHX2
DF	8	ADAMTS15 ADAMTS18 CD34 COL6A1 COL6A2 OLFML2B S100A4 TWIST2
DP	34	ABL1 BMP3 CDH11 CYR61 EPHX2 FGFR1 GABRE GEM HIF3A IZUMO4 LAMC3 LTBP1 LYNX1 MASP1 MRC2 NDNF NOL3 OSR1 PAPPA2 PEMT PRR16 PTCH1 RSPO3 SCUBE2 SMARCD3 SNAI2 SOSTDC1 SPON1 SSC5D TRPS1 UBA7 VCAN WIF1 ZIC4
TAC	10	ATG9B EFNB2 GLI3 KDM2B LEF1 PTCH2 RASGEF1B SLC40A1 TGM1 TLE3
HF-SC	12	ADAMTS17 APOE CRIM1 FOXI3 FOXO6 GPC4 KRT13 LRIG1 PTHLH RGS2 SOX9 TGFB3

DEGs: Differently expressed genes; HF: Hair follicle; Epi: Epidermis; ORS: Outer Root Sheath; Mc: Melanocytes; Mx: Matrix; DF: Dermal Fibroblasts; DP: Total dermal papilla cells; TAC: Transit amplifying cells; HF-SC: Bulge stem cell precursors (Sennett et al., 2015).

The KEGG database was used to refine the potential signaling pathways in our data. The top 20 of the 248 enriched KEGG pathways were selected and presented in **Figure 2C**. Of those, the WNT (19 genes enriched), TGF-β (14 genes enriched) and Hippo (25 genes enriched) signaling pathways were the most promising candidates and were highly correlated to the morphological changes between the selected two developmental stages (**Table 3**). In addition, the Hedgehog signaling pathway (9 genes enriched), which ranked as 21 in pathway enrichment, was another candidate regulating the morphological changes during the stages selected (**Table 3**). This observation is highly consistent with the enrichment of GO terms in positive regulation of cellular and biological processes (**Figure 2B**). Additionally, the enriched gland-related genes in 5 GO terms and 2 KEGG pathway genes were the potential candidates involved in the induction of apocrine sweat glands in coarse wool sheepskin.

Construction of Candidate Gene Interaction Network Functioned in the Elongation of Wool Follicle Pegs and the Induction of Apocrine Sweat Glands in Coarse Wool Sheepskin

Several DEGs were used to construct an mRNA–mRNA interaction network (**Figure 2D**). The genes potentially regulating hair and skin development, gland development, epithelial development (**Tables 1**, **2**) and 4 signaling pathway genes (**Table 3**) were all applied for network construction (**Figure 2D**). The gene *ITGB1* established a small network to regulate the skin dermal fibroblast (*COL6A1* and *COL6A2*) and epidermal development (*FLNB*, *LAMC3*, and *LAMA5*). The gene *KDR*, also named *VEGF*, established the network for regulating epithelial migration and branching. The TGF-β (*BMP7*, *BMPR1A*, *SMAD1*, and *SMAD4*), WNT (*CTNNB1* and *LEF1*) and SHH (*SHH* and *GLI3*) signaling pathways that indicate important regulation of gland and epithelial branching development established the complex networks in the pathway itself and as well as crosstalk among different pathways. For instance, *CTNNB1* was shown to interact with *SMAD1* and *SHH* has connection with *BMP7*, *EGFR*, *FGFR1*, and *LEF1* (**Figure 2D**).

Validation of Potential Candidate Genes Functioned in Wool Follicle Peg Elongation and Apocrine Sweat Gland Induction

A total of 6 genes were selected to evaluate the RNA sequencing results by the qPCR technique. Of those, *BMP7*, *BMPR1A*, *SMAD4*, *WIF1*, and *TGFB3* showed increased expression in the apocrine sweat gland budding stage, while *WNT10B* displayed decreased expression. The expression tendency of these genes is consistent with the RNA sequencing results (**Figure 2E**).

To further explore the enriched candidate genes in our data that were potentially involved in apocrine sweat gland morphogenesis, four antibodies against SOX2 (hair follicle dermal papilla marker), SOX9 (hair follicle stem cell marker), PDGFRB (platelet derived growth factor receptor beta, skin dermal development related) and SMAD5 (BMP signaling) were used to localize the protein expressions during the development of apocrine sweat glands by immunohistochemistry. SOX2 was detected specifically in the dermal condensates (DC) of primary wool follicles in early stages and in the dermal papilla (DP) of well-developed primary wool follicles in later stages (**Figure 3**). The apocrine sweat gland was negative for SOX2 staining (**Figure 3**). It is interesting that SOX2 was only expressed in the DC or DP of primary wool follicles, but not in the secondary wool follicles (**Supplementary Figure S1**). PDGFRB, a cell surface tyrosine kinase receptor, was observed with strong expression in dermal condensates in primary and secondary wool follicles in early stages, but with weak expression in those of well-developed wool follicles (**Figure 4**). The positive staining was also detected in the dermis across the whole development, with weak expression in early stages and strong expression in the upper dermis, especially the area surrounding the wool follicles and gland ducts after the dermal papilla started to form (**Figures 4D,E**). The apocrine sweat glands have been shown negative for PDGFRB staining during all the stages detected. These results suggested that SOX2 and PDGFRB had no direct effect on the induction of sweat gland buds in sheepskin.

The hair follicle bulge stem cell marker *SOX9*, was reported to express in a population of outer root sheath cells (Nowak et al., 2008; Rompolas and Greco, 2014; Purba et al., 2015). In sheepskin, SOX9 was detected with occasional staining in the basal layer of the epidermis and strong signals in the highly

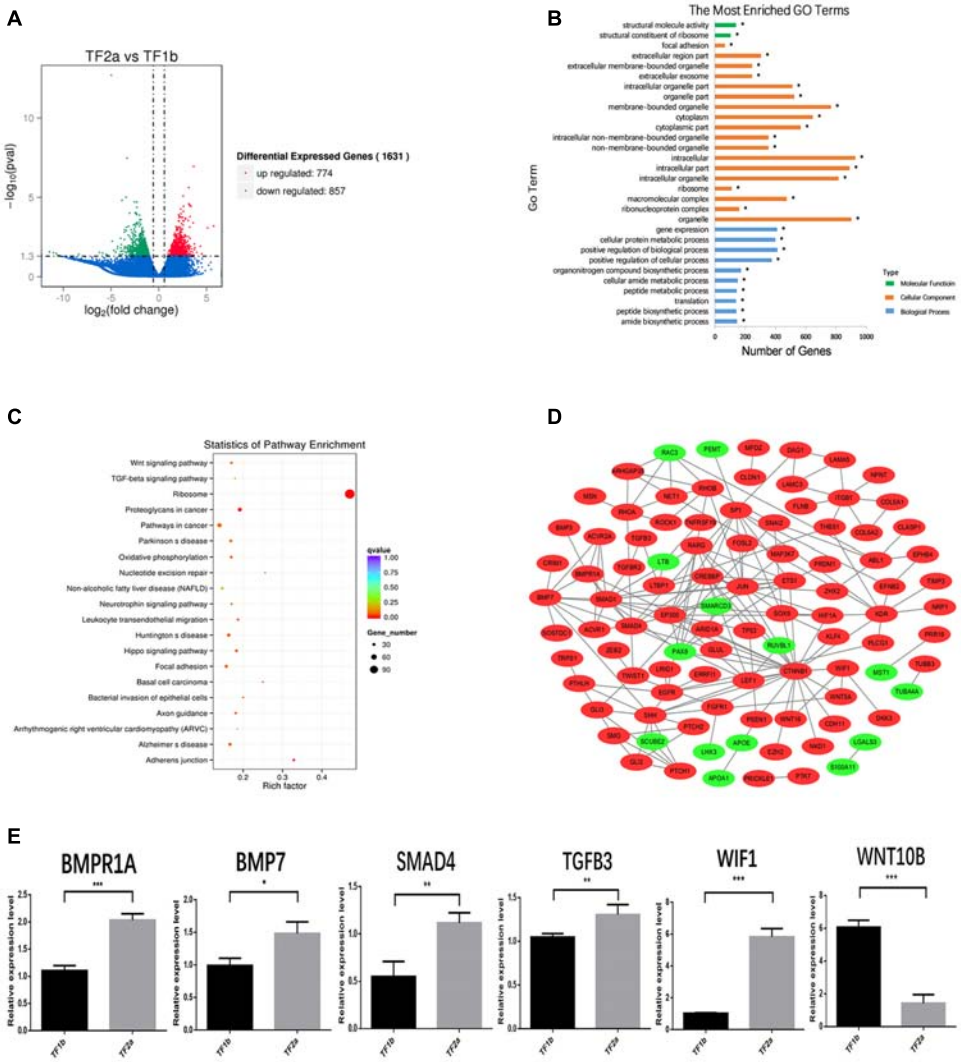


FIGURE 2 | The outcome of bioinformatics analyses of RNA sequencing data of sheepskin with qRT-PCR validation. **(A)** Volcano plot displaying the mRNA transcripts enriched between stage TF2a and stage TF1b of the prenatal sheepskin. The DEGs were enriched to compare the budding stage (TF2a) with the pre-gland stage (TF1b) of apocrine sweat glands in coarse wool sheepskin ($n = 3$) as shown in the volcano plot. There were 1631 differentially expressed transcripts, including 774 upregulated (right, red) and 857 downregulated (left, green), between these two groups. The criteria set up for the enrichment are | log2 (fold change) | > 1 and P value ($P < 0.05$); **(B)** The top 20 GO terms are presented in the enrichment analyses of differentially expressed mRNA transcripts between apocrine sweat gland induction stages (TF2a vs. TF1b) in coarse wool sheepskin. A total of 786 terms were significantly enriched ($P < 0.05$) in the categories of biological process (blue), molecular function (green), and cellular components (yellow red). **(C)** The top 20 KEGG pathways are displayed in the enrichment analyses of differentially expressed mRNA transcripts in apocrine sweat gland stages (TF2a vs. TF1b) in coarse wool sheepskin. The top 20 out of 248 terms of differently expressed mRNA transcripts were grouped and displayed ($P < 0.05$). Of those, three signaling pathways (WNT, TGF-β, and Hippo signaling pathways), focal adhesion and adherent junctions were potentially involved in histological changes during the morphogenesis of apocrine sweat gland (TF2a vs. TF1b) in coarse wool sheepskin. **(D)** The mRNA–mRNA interaction networks were constructed by using the potential candidate genes involved in the development of skin, wool follicles and glands. The WNT, TGF-beta, and SHH signaling pathways were clearly grouped. Another two groups networked by KDR and ITGB1 were potentially involved in basement membrane and cell proliferation. **(E)** A total of 6 DEGs were selected for quantitative real-time PCR (qRT-PCR) validation. The expression patterns of *BMP1A*, *BMP7*, *SMAD4*, *TGFB3*, *WIF1*, and *WNT10B* are consistent with the tendency of mRNA sequencing results by using the $2^{-\Delta\Delta Ct}$ method and GAPDH as internal control. Data are presented as mean ± SD ($n \geq 3$). $^{*}P < 0.05$, $^{**}P < 0.01$ $^{***}P < 0.001$ (Student's *t*-test).

proliferated and differentiated epidermal compartments of wool follicles during the early and later stages (**Figure 5**). Strong SOX9 signals were also detected in the apocrine sweat glands, from the precursor cell patches, to the budded and elongated apocrine sweat gland ducts (**Figures 5B–E**). More details showed that SOX9 was first detected with weak expression in the cell aggregates that indicated the precursor/progenitor cells of apocrine sweat glands (**Figure 5B**). Strong expression of SOX9 was then observed initially in budding sites, later in the germs and the elongated ducts of the apocrine sweat glands (**Figures 5C–E**).

Immunohistochemistry of pSMAD5 showed broad expression in sheepskin across the developmental stages. The epidermis

TABLE 2 | The GO terms are specifically enriched in skin and hair follicle development, epithelial development and gland development.

Description	Number	Genes
Salivary gland morphogenesis and development	8	BMP7 DAG1 EGFR FGFR1 LAMA5 NRP1 SNAI2 TGFB3
Branching involved in salivary gland morphogenesis	6	BMP7 DAG1 FGFR1 LAMA5 NRP1 SNAI2
mammary gland epithelial cell differentiation	4	HIF1A, SMO, PTCH1, LBH
Gland development	42	ABL1 **APOA1** ARHGAP35 ARID5B BMP7 BMPR1A CTNNB1 **COBL** DAG1 EGFR EZH2 FGFR1 GLI2 GLI3 HIF1A JUN LAMA5 LBH LEF1 **LHX3 MKI67** MSN **MST1** NRP1 **PAX8** PSEN1 PTCH1 RARG **RHBDD3** SMAD4 SMO SNAI2 SOCS2 SOSTDC1 SOX9 SP3 SULF2 **TAF10** TGFB3 TGFBR2 WLS WNT5A
Gland morphogenesis	20	BMP7 DAG1 EGFR FBR2 FGFR1 GLI2 GLI3 LAMA5 **MKI67** MSN **MST1** NRP1 PTCH1 RARG SNAI2 SOSTDC1 SOX9 SULF2 TGFB3 WNT5A
Hair follicle development	14	APCDD1 CTNNB1 EGFR LAMA5 LGR4 LGR5 LHX2 LRP4 SMAD4 SMO SOSTDC1 SOX9 TNFRSF19 WNT5A
Skin development	33	**ALOX12B** APCDD1 ARRDC3 CLDN1 CTNNB1 CYP26B1 **DACT2** EGFR ERRFI1 FLNB FOSL2 FRAS1 H2AFY2 LAMA5 LGR4 LGR5 LHX2 LRP4 **LTB** PSEN1 PTCH2 PTGES3 ROCK1 SMAD4 SMO SOSTDC1 SOX9 TNFRSF19 **TGM1** WDR48 WNT16 WNT5A ZFP36L1
Embryonic epithelial tube formation	23	ABL1 ARHGAP35 ARID1A BMP7 **COBL** CTNNB1 HIF1A KDM2B LHX2 LUZP1 OSR1 **PAX8** PRICKLE1 PTCH1 PTK7 RARG RGMA SEMA4C SOX9 TMED2 TWIST1 WNT5A ZEB2
Epithelial cell migration	27	ADGRA2 **APOA1 APOE** ARSB EMP2 EFNB2 ETS1 EPHB4 CLASP1 FGFR1 HIF1A ITGB1 JUN KDR KLF4 LOXL2 MYH9 NRP1 PLCG1 **RAB25** RHOB **S100A2** SOX9 TGFBR2 THBS1 WNT5A ZEB2
Branching morphogenesis of an epithelial tube	21	ACVR1 BMP7 CTNNB1 DAG1 DLG5 EDNRA GLI GLI3 LAMA5 LEF1 LGR4 MMP14 NPNT NRP1 **PAX8** PTCH1 SMAD4 SMO SOX9 TGFBR2 WNT5A
Morphogenesis of a branching epithelium	24	ACVR1 BMP7 CTNNB1 DAG1 DLG5 EDNRA FGFR1 GLI2 GLI3 LAMA5 LEF1 LGR4 MMP14 NPNT NRP1 **PAX8** PTCH1 RSPO3 SMAD4 SMO SNAI2 SOX9 TGFBR2 WNT5A
Morphogenesis of a branching structure	25	ACVR1 BMP7 CTNNB1 DAG1 DLG5 EDNRA FGFR1 GLI2 GLI3 LAMA5 LEF1 LGR4 MMP14 NPNT NRP1 **PAX8** PRDM1 PTCH1 RSPO3 SMAD4 SMO SNAI2 SOX9 TGFBR2 WNT5A
Morphogenesis of an epithelial fold	8	BMP7 CTNNB1 EGFR GLI2 HIF1A LUZP1 SOSTDC1 WNT5A
Morphogenesis of an epithelial bud	5	BMP7 CTNNB1 GLI2 SOSTDC1 WNT5A

The bold black font represents the downregulated DEGs.

TABLE 3 | The DEGs enriched in three signaling pathways are potentially related to sweat gland development.

Signaling Pathway	Number	Genes
WNT	19	CREBBP CTNNB1 EP300 GPC4 JUN LEF1 MAP3K7 NKD1 PRICKLE1 PSEN1 RAC3 RHOA **RUVBL1** SMAD4 TP53 WIF1 **WNT10B** WNT16 WNT5A
TGF-beta	14	ACVR1 ACVR2A BMP7 BMPR1A CREBBP EP300 RHOA ROCK1 SMAD1 SMAD4 SP1 TGFB3 TGFBR2 THBS1
Hedgehog	9	GLI2 GLI3 PTCH1 PTCH2 SHH SMO **WNT10B** WNT16 WNT5A
Hippo	25	ACTB ACTG1 BMP7 BMPR1A CCND1 CTNNA1 CTNNB1 FRMD6 GLI2 LEF1 PPP1CB PPP1CC SMAD1 SMAD4 SNAI2 TGFB3 TGFBR2 **WNT10B** WNT16 WNT5A WWTR1 YWHAE YWHAG YWHAQ YWHAZ

The bold black font represents the downregulated DEGs.

and dermis in addition to the wool follicles and sweat glands were all positive for pSMAD5 antibody staining (**Figure 6**). Detailed inspection revealed that the strongest positive signals of pSMAD5 were observed in the basal layer of the epidermis and the epidermal compartments of wool follicles during all the stages detected. During the apocrine sweat gland development, pSMAD5 antibody was localized initially in the few pre-gland precursor/progenitor cells with weak expression (**Figures 6B,C**), and then in the budded gland loops and later in the elongated gland ductal portions with strong expression (**Figures 6D,E**).

DISCUSSION

Chicken and mouse are the most widely used models to study the mechanisms underlying skin and feather/hair follicle morphogenesis, development, cycling and regeneration. Though lots of genetically modified mouse/chicken models were created to perform the functional study of individual or combined genes or gene networks in skin and feather/hair research, there are still lots of questions that remain unclear due to the limitation of animal models themselves. Chicken skin consists of only feather follicles with no sweat glands or sebaceous glands (**Figure 7**). Hair follicles and sebaceous glands are the two appendages that exist broadly in mouse dorsal skin, while eccrine sweat glands as another appendage remain specifically in mouse paw skin (**Figure 7**). This additional and regional localization of eccrine sweat glands in footpad skin is a good

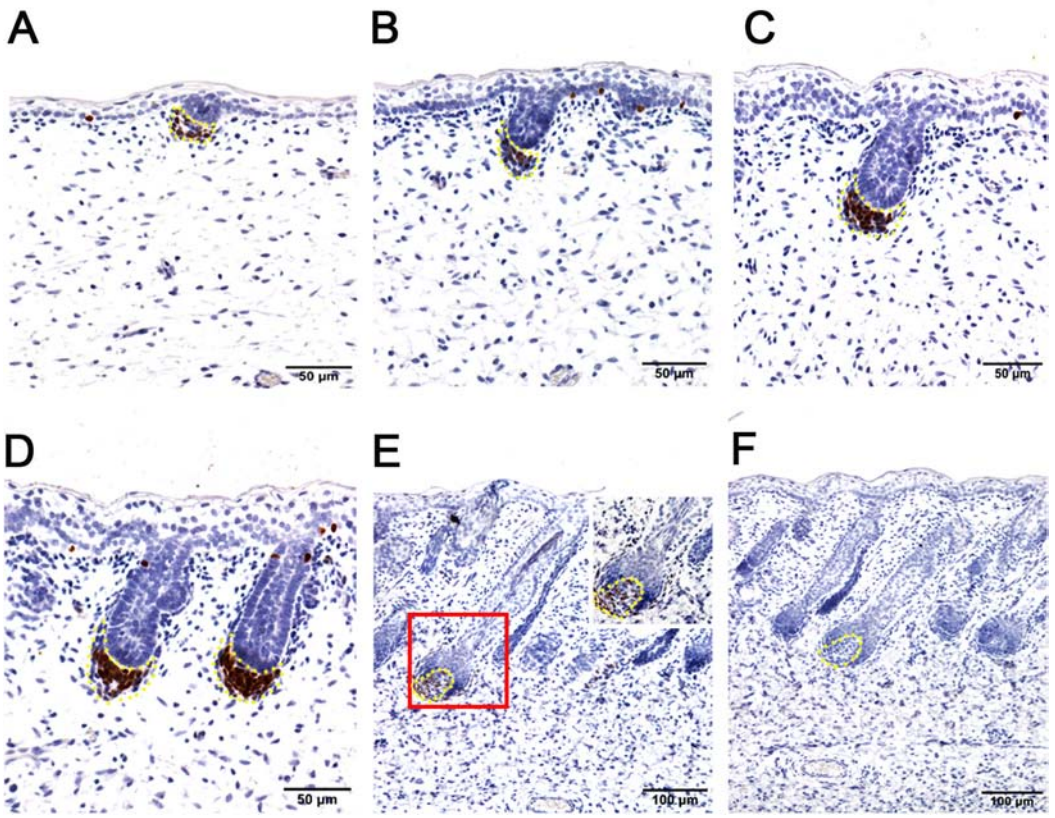


FIGURE 3 | Hair follicle marker SOX2 is negative in apocrine sweat glands in prenatal sheepskin as detected by immunohistochemistry. SOX2 is specifically localized in the dermal condensates/dermal papilla of primary wool follicles, not in those of secondary wool follicles and apocrine sweat glands. **(A–E)** The strong positive signals of SOX2 antibody staining are restricted in the dermal condensates associated with the primary wool follicle placodes **(A–D)** in early stages and in the dermal papilla in later stage **(E)**. The secondary wool follicles and apocrine sweat glands are all negative for SOX2 immunohistochemistry. **(F)** The negative control is displayed without applying the primary antibody in immunohistochemistry. **A–D** Bar, 50 μm. **E** and **F** Bar, 100 μm.

point to study the subtype appendage determination during the early development of skin. The hair follicles in mouse dorsal skin display three synchronized developmental waves prenatally and cyclic growth pattern postnatally. Moreover, numerous genetic modified mouse models were generated in recent years, benefiting from the profoundly developed genome editing technique. All the advantages contribute to making the mouse model a broadly used system for deep understanding of the skin biology.

Human skin is distinctive from other animals partially in that the sweat glands exist across the skin and the subtypes of sweat glands are either eccrine sweat glands on the non-hairy area or apocrine sweat glands on the hairy area (**Figure 7**). Hence the absence of apocrine sweat glands in mouse and chicken skin restrains the further understanding of the complexity of human skin biology and skin diseases, like armpit and body odor. Sheep could serve as an additional system to further explore the knowledge of apocrine sweat glands since sheepskin has sweat glands that are similar to those located in human armpits (**Figure 7**). Until now, there has been good understanding of the development of eccrine sweat glands (Klaka et al., 2017; Kurata et al., 2017), but not that of apocrine sweat glands. Our study is the first report revealing the complex molecular network regulating early development, especially the morphogenesis of apocrine sweat glands using sheep as a model. Coarse wool sheep develop primary and secondary wool follicles that are similar to the generation of hair follicles in mouse dorsal skin. The occurrence of apocrine sweat glands is potentially in accompaniment with the generation of secondary wool follicles as indicated in **Figure 1C**. The observation that the eccrine sweat glands were initiated to form the first wave of pre-germ at E16.5 (the period of secondary follicles emergence) in the proximal footpad and later at E17.5 in the distal footpad (Schlake, 2007; Cui et al., 2014) implied that the two types of sweat glands shared the similar induction time schedule during the early morphogenesis.

At the induction stage, a few cells packing together on the lateral side of the half length of primary wool follicle germ indicated the location of precursor/progenitor cells of apocrine sweat glands approximately at embryonic day 75 (**Figure 1C**). This unilateral pattern formation is different from the bilateral pattern of sebaceous glands that develop after the apocrine glands in sheepskin. It is also different from the *de novo* pattern formation of eccrine sweat glands that develop from the crosstalk of epidermal and dermal layers of the skin (Lu et al., 2016).

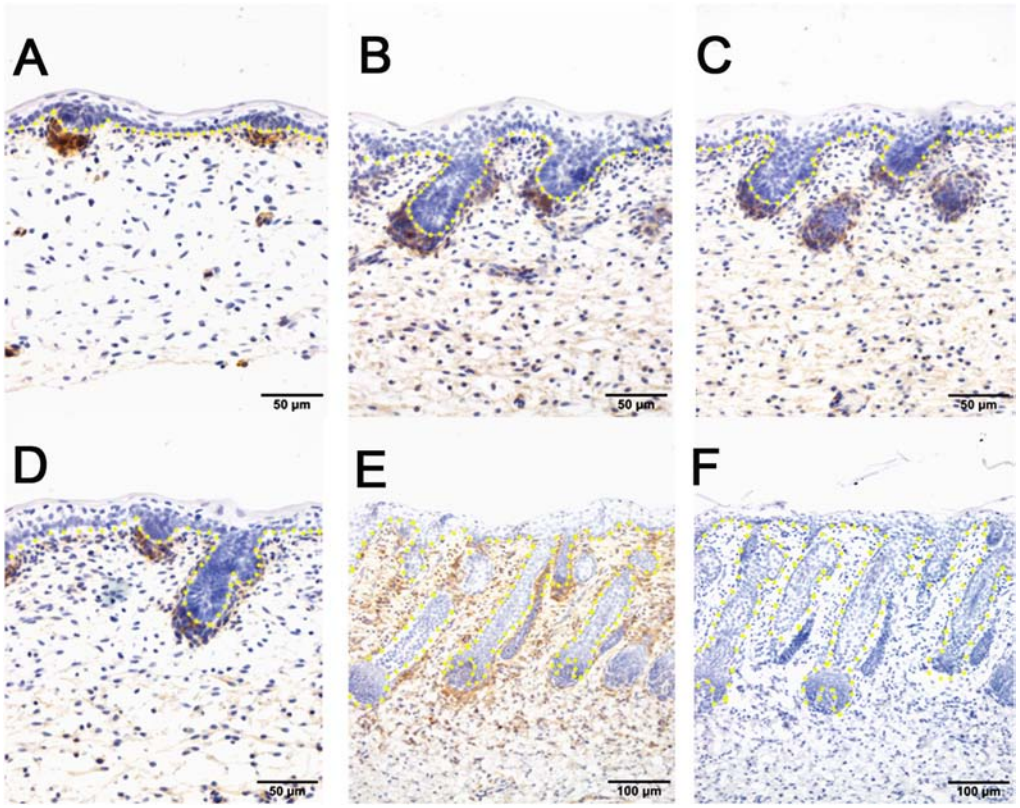


FIGURE 4 | The localization of PDGFRB by immunohistochemistry is positive in dermal-originated cells and negative in apocrine sweat glands in the prenatal sheepskin. **(A)** PDGFRB is highly expressed in the dermal condensates of primary wool follicles and occasionally in dermal fibroblast before apocrine sweat gland induction; **(B–D)** PDGFRB is expressed in dermal condensates of primary and secondary wool follicles, and not in the precursor cells or the bud of apocrine sweat glands; **(E)** Strong PDGFRB expression is detected in the upper dermis of the sheepskin, especially the location surrounding the wool follicles. PDGFRB is surprisingly negative in the dermal papilla of the well-developed wool follicles and also negative in the apocrine sweat glands. **(F)** The negative control is displayed without applying the primary antibody in immunohistochemistry. **A–D** Bar, 50 μm. **E** and **F** Bar, 100 μm.

These compact cell patches gradually grew outward from the adjacent outer root sheaths of primary wool follicles to form the short and later long ductal bud as indicated by the H&E stain in our study (**Figure 1**). Then the bud extended closely parallel to the primary wool follicle peg and developed the secretory portions to become mature apocrine sweat glands in later stages.

The two developmental stages applied in the RNA-sequencing program represent the pre-gland phase (stage TF1b, **Figure 1B**) and gland bud phase (stage TF2a, **Figure 1E**) of apocrine sweat glands. At the selected gland budding stage, the primary wool follicles grew downward to the dermis to develop follicle pegs and at the same time secondary wool follicles started to generate the placode and associated dermal condensates. Briefly, the prominent structural changes between the two stages are the thickened epidermis and dermis, elongated wool follicle pegs, enlarged dermal condensates, initiated secondary wool follicle placodes and emerged apocrine sweat gland germs. By analyses of 1631 DEGs, a series of genes was enriched to function in skin development (33 genes) and hair follicle development (14 genes) (**Table 2**). These genes are potentially responsible for the morphological changes for skin epidermal and dermal thickening as well as wool follicle germ elongation. Most of the genes showed the increased expression trends in line with the positive regulation of cellular processes and biological processes in enriched GO terms. Further analyses showed that our data were highly comparable with those of genes regulating the hair follicle development in P5 mouse skin. The overlapped genes between these two groups covered regulatory genes responsible for all the compartments of the skin and hair/wool follicles. Genes regulating each compartment of wool/hair follicles were analyzed and recorded in **Table 1** and represented by *TGFB1* and *WNT16* for epidermal development, *ADAMTS15* and *COL6A1* for the dermal fibroblast development, *LAMA5* and *SOX9* for outer root sheath development, *LRIG1* and *SOX9* for hair follicle stem cell development. These results indicate that the epidermal part of wool follicles was rapidly growing during the two stages we detected. The most promising result is the enrichment of 34 genes for wool follicle dermal papilla development represented by the commonly used dermal condensate marker genes *BMP3*, *SOSTDC1*, *TRPS1* and *WIF1*. These analyses were consistent with the enlarged dermal condensates associated with the primary follicle pegs and the newly formed secondary wool follicles as shown in **Figure 1C**.

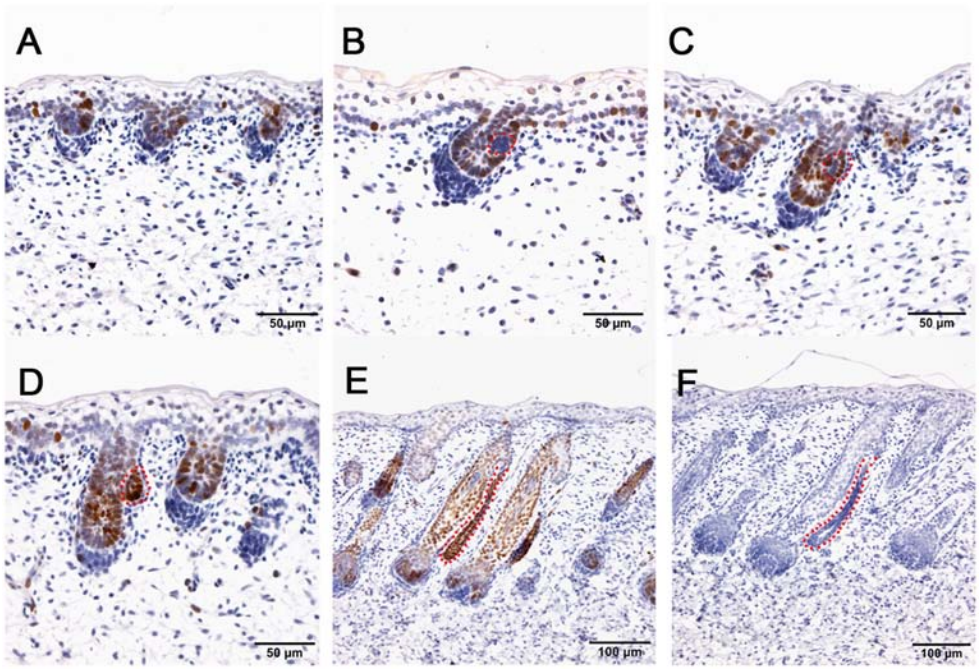


FIGURE 5 | SOX9 is localized in the wool follicle pegs and apocrine sweat glands (red dashed line). **(A)** SOX9 is expressed in the placodes of primary wool follicles and occasionally in the basal layer; **(B)** As the primary wool follicles grow downward into the dermis, SOX9 is strongly expressed in the primary wool follicle pegs and basal layer, and weakly in the precursor cells of the apocrine sweat glands; **(C–D)** SOX9 is expressed in the apocrine sweat gland budding point **(C)** and buds **(D)**, and also in wool follicle pegs. **(E)** SOX9 is strongly expressed in the ducts of apocrine sweat glands. It is also expressed in the elongated epidermal compartments of wool follicles. **(F)** The negative control is displayed without applying the primary antibody in immunohistochemistry. **A–D** Bar, 50 μm. **E** and **F** Bar, 100 μm.

Though these two datasets do not originate from the same developmental stages, they do share similar gene networks that govern the skin and hair/wool follicle development between mouse and sheep. It indicates that the pre-mature wool follicles in the selected stages of sheepskin and mature hair follicles in P5 mouse skin were regulated by partially conserved candidate genes with different dosages or locations. The overlapping marker genes in **Table 1** and the immunohistochemistry of selected candidates in **Figures 3**–**6** clearly stated this notion.

SOX2, a hair follicle maker, was specifically positive in the dermal condensates/dermal papilla of the primary wool follicles and surprisingly negative in the secondary wool follicles (**Figure 3** and **Supplementary Figure S1**). This expression pattern is partially different from those of the mice in that Sox2 was detected in both primary and secondary hair follicles, not in the third wave zigzag follicles in mouse dorsal skin (Graham et al., 2003; Driskell et al., 2009). PDGFRB is one of the types of PDGF receptors which can mediate the biological actions of PDGF and is related to the development of many organs (Claesson-Welsh et al., 1988; Gronwald et al., 1988; Mellgren et al., 2008). Moreover, Pdgfrb was expressed in the dermis and dermal condensates of E14.5 mouse skin (Rezza et al., 2015) and disruption of *Pdgfrb* signaling impaired proliferation and dermal fibroblast migration (Gao et al., 2005; Rajkumar et al., 2006). In sheepskin, PDGFRB displayed focal expression in the dermal papilla of the pre-mature wool follicles and the fibroblast in early stages, and later in the upper part of the mesenchyme, especially the area surrounding the wool follicles. The signals of SOX2 and PDGFRB were both absent from the ductal buds of the apocrine sweat glands, indicating these two genes were not important for the early morphogenesis of apocrine sweat glands. And it also suggests that the regulatory networks of dermal-originated hair/wool follicle compartments were different from those of apocrine sweat glands branched from the epidermal-originated outer root sheaths of the wool follicles. Sox9 was reported to mainly express in the outer root sheath and the bulge of hair follicles (Nowak et al., 2008; Rompolas and Greco, 2014; Purba et al., 2015). The detection of SOX9-positive signals in epidermal compartments of the wool follicles, the inter-follicular basal layers and the apocrine sweat gland ducts implied that hair/wool follicles and apocrine sweat glands partially share some key regulators, especially the molecules regulating the outer root sheaths during the morphogenesis. The expression of SOX9 was initially detected in the precursor cells (the cell aggregates located on the lateral side of the wool follicle peg) of the apocrine sweat gland and later in the branched gland germs and straight gland ducts (**Figure 5**). The aggregate precursor cells marked the initiation of apocrine sweat glands. Then these few cells proliferate, differentiate and migrate to the edge of the follicle peg and form the gland cavity with the small opening to the upper part of the outer root sheath.

The induction of apocrine sweat gland germs and the elongated epidermal compartment of wool follicle pegs are highly

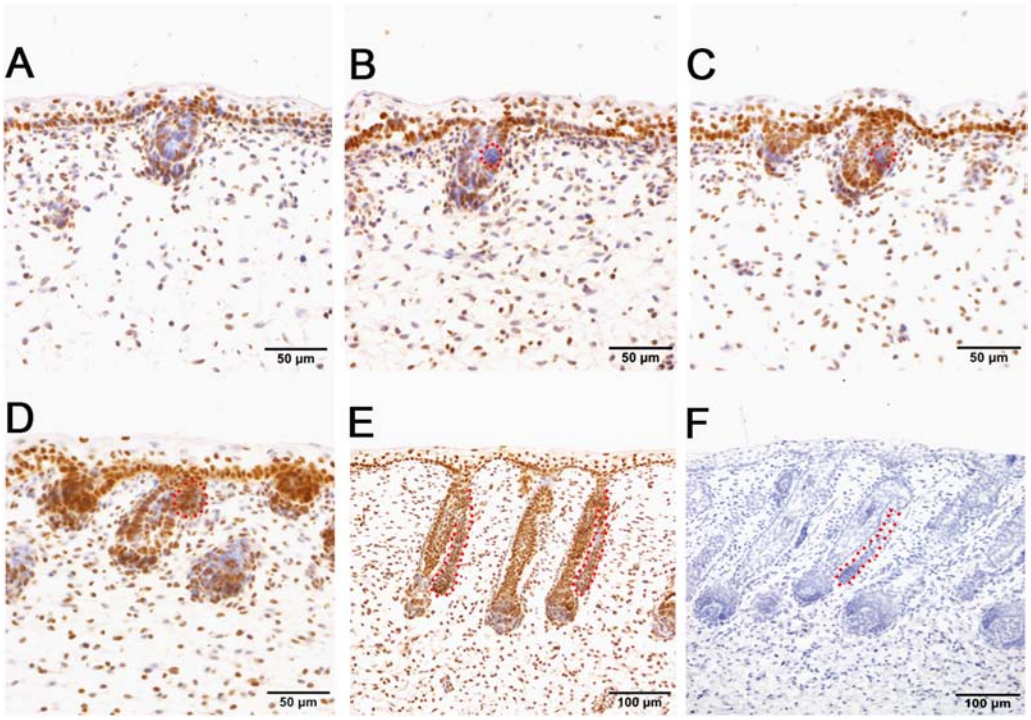


FIGURE 6 | pSMAD5 is broadly expressed in the compartments of prenatal sheepskin. **(A)** pSMAD5 is expressed in the dermis, epidermis and primary wool follicle placodes and associated dermal condensates. **(B–E)** pSMAD5 is expressed in the precursor cells, buds and ducts of apocrine sweat glands and is also expressed in all skin compartments. **(F)** The negative control is displayed without applying the primary antibody in immunohistochemistry. **A–D** Bar, 50 μm. **E** and **F** Bar, 100 μm.

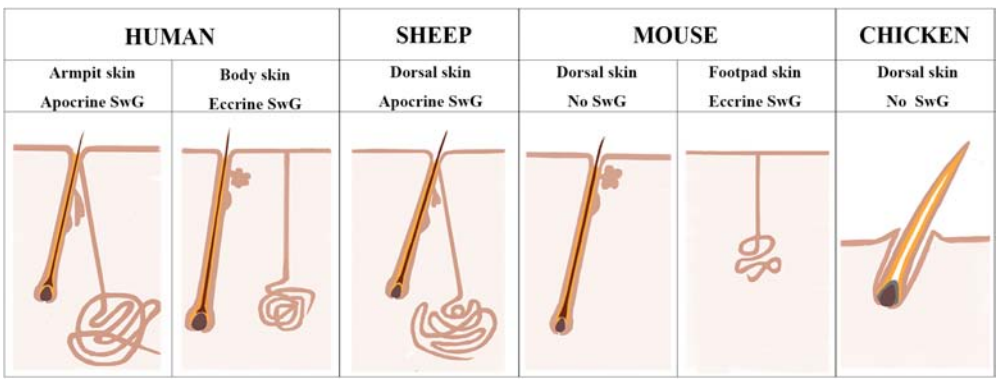


FIGURE 7 | Sheepskin is an additional system to study human apocrine sweat glands. A diagram is drawn up to compare the complexity of sweat glands in human skin with those in the widely used animal models (mouse and chicken) and also in sheep used in current study. There are two types of sweat glands that exist in human skin. Eccrine sweat glands are distributed in hairless skin regions, while apocrine sweat glands are exhibited in the armpit skin. Sheep dorsal skin contains apocrine sweat glands that are similar with those of human armpit skin. Mouse dorsal skin has no sweat glands. Eccrine sweat glands exist exclusively in mouse footpad skin. Chicken skin is completely absent of sweat glands.

correlated with the epithelial cell migration, differentiation, and morphogenesis of epithelial branching or tube formation (**Table 2**). The enrichment of 7 categories of genes involved in epithelial branching or tube formation is consistent with the observation that the wool follicle is tube-like in structure, branching from the skin basal layer, and the apocrine sweat glands protrude from the outer root sheath to form the branching with two layers (basal layer and supra-basal layer) surrounding the cavity of the gland (**Figure 1I**). These structures mostly originate from the epithelia in line with the enrichment of epithelial-related GO terms and candidate genes.

The most interesting point is the enrichment of 5 gland-related GO terms in our data. A series of genes represented by *BMP7*, *FGFR1*, *GLI2*, *LAMA5*, *SOX9*, and *WNT5A* enriched in gland morphogenesis (18 genes increased and 2 genes decreased) and gland development (34 genes increased and 8 genes decreased)

were reported to be involved in general gland development, including the salivary gland, mammary gland and prostate gland development. Though the apocrine sweat gland in our study is structurally different from those types of glands mentioned above, they do partially share the regulatory genes functioned in the early morphogenesis, especially the ductal formation. A total of 8 genes (*BMP7*, *DAG1*, *EGFR*, *FGFR1*, *LAMA5*, *NRP1*, *SNAI2*, and *TGFB3*) were grouped specifically for salivary gland morphogenesis and development, while 6 genes (*BMP7*, *DAG1*, *FGFR1*, *LAMA5*, *NRP1*, and *SNAI2*) were enriched for branching involved in salivary gland morphogenesis. Recently, a report suggested that conditional deletion of *Nrp1* in mammary epithelial cells delayed mammary development, particularly the ductal extension (Liu et al., 2017). Lama5 and Dag1 were broadly expressed in the basement membrane of skin and hair follicles in mouse models to maintain skin integrity. Previous reports showed that *Lama5* played important roles in hair peg elongation and skin homeostasis since conditional knockout mice *Lama5*^{Ker5} showed delayed hair growth in early age, abnormal follicle down-growth and decreased hair follicle density in adult animals (Wegner et al., 2016). The fact that LAMA5 was highly expressed in the basement membrane of straight ducts and secretory portions of the human eccrine sweat glands (Kurata et al., 2017) implied that the increased expression of *LAMA5* in the apocrine sweat gland budding stage was responsible for both the wool follicle peg elongation and the ductal formation of the apocrine sweat glands. *SNAI2* (*Slug*) was shown to determine the mammary stem cells in line with Sox9 (Guo et al., 2012). The upregulation of *SNAI2* in our data indicates that *SNAI2* was involved in apocrine sweat gland induction. These analyses further proved the notion that the molecular networks controlling gland morphogenesis were partially shared among diverse glands (salivary gland, mammary gland and apocrine sweat gland) and functionally different from those of hair/wool follicles development.

In addition to the GO terms mentioned above, we also significantly enriched three signaling pathways (WNT, TGF-β, and Hedgehog) that may be involved in apocrine sweat gland morphogenesis and development. In our data, we accumulated 19 genes of the WNT signaling pathway. The WNT pathway was indispensable for induction and development of hair follicles and eccrine sweat glands (Chen et al., 2012; Cui et al., 2014). Among these genes, *Wnt5a* was involved in the proper development of bud outgrowth and branching point formation of the prostatic gland (Huang et al., 2009b). The application of Wnt5a protein in a tissue culture system also inhibited the ductal branching and extension of mammary gland (Roarty and Serra, 2007). During this process, *Wnt5a* was supposed to function as a downstream effector of TGF-β signaling that showed similar regulatory impact on mammary gland development (Roarty and Serra, 2007). The upregulation of *WNT5A* in our data indicated that *WNT5A* is involved in early development of apocrine sweat gland. But whether or not *WNT5A* inhibits the formation of apocrine sweat gland duct requires further study. *BMP7*, *BMPRIA*, *SMAD1*, and *SMAD4* were enriched in the TGF-β signaling pathway. *Smad1* was focally expressed in the dermis of eccrine sweat gland germs during the induction stage (O'Shaughnessy et al., 2004). *Smad4* is a common partner interacting with *Smad1*, *Smad5*, and *Smad8* to mediate BMP signaling, and with *Smad2* and *Smad3* to mediate TGF-β subfamily signaling (ten Dijke and Hill, 2004). *Smad4* conditional knockout mice exhibited abnormal proliferation and differentiation, particularly with increased cell proliferation in the outer root sheaths and epidermis, mainly due to the blockage of TGF-β subfamily (Owens et al., 2008). Several mouse models reported that upregulation of BMP signals was important for eccrine sweat gland development. The overexpression of *Noggin* (antagonist of BMP) in *K14-Noggin* transgenic mice displayed increased hair follicle density in body skin, and transformed eccrine sweat glands in footpads into hair follicles (Plikus et al., 2004). The *Bmpr1a* conditional knockout mice converted the eccrine glandular appendage fate to form hair follicle-like structures in mouse footpad skin (Lu et al., 2016). These mouse models related to the BMP pathway highly suggest that upregulation of BMP signals favor eccrine sweat gland development. *Bmp7*, a conserved secreted molecule of BMP family, was detected in both epidermal and dermal layers in mouse E13.5 skin, broadly expressed in the epithelium of the salivary gland, the immature hair follicles (inner and outer root sheaths, hair shaft and dermal papilla) and the mesenchyme surrounding the hair follicles, and highly enriched in mature hair follicles (dermal papilla and outer root sheaths) (Zouvelou et al., 2009). The *Bmp7* conditional knockout mice showed abnormal hair follicles with enlarged root sheaths. *BMP7* also proved to regulate the branching of lacrimal glands and prostate glands (Dean et al., 2004). The increased expression of *BMP7* in our data, combined with the reported functional study of *BMP7* in gland development and the expression pattern of *BMP7* in the outer root sheath of hair follicles, strongly suggest that *BMP7* was potentially involved in the morphogenesis of apocrine sweat glands. In our study, the localization of pSmad5 was observed at the branching point and the germs of the apocrine sweat glands with strong expression (**Figure 6**). This expression pattern of pSmad5 suggested that BMP signaling was activated at the gland induction site as that of the eccrine sweat gland. Hence, the enrichment of TGF-β, particularly BMP-signaling genes *BMP7*, *BMPRIA*, *SMAD1*, and *SMAD4* in our sequencing analysis, combined with the mouse models of the BMP pathway discussed above, highly suggest that BMP signaling is a positive regulator of apocrine sweat gland induction.

The competence of high-BMP and low-SHH signals in a short developmental period established the initiation of eccrine sweat glands instead of the hair follicles in mouse ventral foot pad skin as discussed previously (Lu et al., 2016). Interestingly, the SHH signaling pathway was enriched in our data as well, including genes *SHH*, *GLI2*, and *GLI3*. The SHH signaling pathway was shown to positively regulate the down-growth of hair follicle pegs and negatively determine the induction of eccrine sweat glands in a short developmental stage (Cui et al., 2011, 2014; Lu et al., 2016). The increased expression of SHH signaling in our data was potentially responsible for wool follicle elongation and gland duct extension. It is also possible that SHH regulated the induction of apocrine sweat glands at the earlier stage and at the restricted branching point.

The histological study in our current report clearly stated that apocrine sweat glands in sheepskin were branched from the outer root sheaths of the primary wool follicles. A total of 43 genes enriched in 5 categories of gland morphogenesis and development in our data, implying that the regulatory network for the morphogenesis of apocrine sweat glands in sheepskin was partially conserved with the other glands, particularly mammary glands, salivary glands and eccrine sweat glands, though the originations of apocrine sweat glands and eccrine sweat glands are different. Of those, the BMP and WNT signaling pathway genes (*BMP7*, *BMPR1A*, *SMAD1*, *SMAD4*, and *WNT5A*) and the 8 gland-related genes are the most promising candidates potentially exhibiting positive regulation of apocrine sweat gland induction. The negative regulators during this process are not specified in our data. It may be that SHH pathway genes (*SHH*, *GLI1*, and *GLI2*) functioned in the branching point during the induction of apocrine sweat glands. Until now, few studies have been conducted on apocrine sweat gland development. Moreover, transcriptome studies of sheepskin have also been widely reported, but little attention has been paid to the development of apocrine sweat glands. Our report is the first to reveal the complex molecular network interaction in the induction stage of apocrine sweat glands in coarse wool sheepskin and will contribute to the better understanding of the histology, physiology and pathology of apocrine sweat glands and associated diseases in humans and companion animals.

DATA AVAILABILITY STATEMENT

The RNA-seq data were submitted to the NCBI database under the SRA Accession: PRJNA507468.

AUTHOR CONTRIBUTIONS

CM designed the experiment, wrote and revised the manuscript. SL wrote the manuscript, analyzed the data, and performed the qRT-PCR and immunohistochemistry. XZ, YN, and MY involved in sample collection and staining. WC, YT, and XH adjusted the picture format. ZL, YH, HQ, QQ, QP, and DC participated in the collection of samples.

FUNDING

This work was supported by the grants from the National Key R&D Program of China (2018YFD0501301) and Special Funds for Basic Research Projects in Central Universities (2662015PY007).

ACKNOWLEDGMENTS

We thank Dr. Liqing Ma and Mr. Xiaoqiang Zhang, and the Animal Husbandry and Veterinary Station for providing experimental equipment. And we also thank all the staff at the local abattoir for their great help during the sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00739/full#supplementary-material

REFERENCES

Baharak, A., Reza, K., Shahriar, D., Omid, A., Daruoosh, V., and Nasrin, A. (2012). Metastatic apocrine sweat gland adenocarcinoma in a terrier dog. *Asian Pac. J. Trop. Biomed.* 2, 670–672. doi: 10.1016/S2221-1691(12)60118-X

Chen, D., Jarrell, A., Guo, C., Lang, R., and Atit, R. (2012). Dermal beta-catenin activity in response to epidermal Wnt ligands is required for fibroblast proliferation and hair follicle initiation. *Development* 139, 1522–1533. doi: 10.1242/dev.076463

Claesson-Welsh, L., Eriksson, A., Moren, A., Severinsson, L., Ek, B., Ostman, A., et al. (1988). cDNA cloning and expression of a human platelet-derived growth factor (PDGF) receptor specific for B-chain-containing PDGF molecules. *Mol. Cell. Biol.* 8, 3476–3486. doi: 10.1128/MCB.8.8.3476

Cui, C. Y., Kunisada, M., Childress, V., Michel, M., and Schlessinger, D. (2011). Shh is required for Tabby hair follicle development. *Cell Cycle* 10, 3379–3386. doi: 10.4161/cc.10.19.17669

Cui, C. Y., Yin, M., Sima, J., Childress, V., Michel, M., Piao, Y., et al. (2014). Involvement of Wnt, Eda and Shh at defined stages of sweat gland development. *Development* 141, 3752–3760. doi: 10.1242/dev.109231

Dean, C., Ito, M., Makarenkova, H. P., Faber, S. C., and Lang, R. A. (2004). Bmp7 regulates branching morphogenesis of the lacrimal gland by promoting mesenchymal proliferation and condensation. *Development* 131, 4155–4165. doi: 10.1242/dev.01285

Driskell, R. R., Giangreco, A., Jensen, K. B., Mulder, K. W., and Watt, F. M. (2009). Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis. *Development* 136, 2815–2823. doi: 10.1242/dev.038620

Fujiwara-Igarashi, A., Shimizu, K., Michishita, M., Yu, Y., Hamamoto, Y., Hasegawa, D., et al. (2017). A cat with suspected laryngeal metastasis with mucosal irregularity resulting from apocrine/salivary gland adenocarcinoma in the head. *J. Vet. Med. Sci.* 79, 1916–1919. doi: 10.1292/jvms.17-0242

Gao, W., Sun, W., Yin, J., Lv, X., Bao, J., Yu, J., et al. (2017). Screening candidate microRNAs (miRNAs) in different lambskin hair follicles in Hu sheep. *PLoS One* 12:e0176532. doi: 10.1371/journal.pone.0176532

Gao, Z., Sasaoka, T., Fujimori, T., Oya, T., Ishii, Y., Sabit, H., et al. (2005). Deletion of the PDGFR-beta gene affects key fibroblast functions important for wound healing. *J. Biol. Chem.* 280, 9375–9389. doi: 10.1074/jbc.M413081200

Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. *Neuron* 39, 749–765. doi: 10.1016/S0896-6273(03)00497-5

Gronwald, R. G., Grant, F. J., Haldeman, B. A., Hart, C. E., O'hara, P. J., Hagen, F. S., et al. (1988). Cloning and expression of a cDNA coding for the human platelet-derived growth factor receptor: evidence for more than one receptor class. *Proc. Natl. Acad. Sci. U.S.A.* 85, 3435–3439. doi: 10.1073/pnas.85.10.3435

Guo, W., Keckesova, Z., Donaher, J. L., Shibue, T., Tischler, V., Reinhardt, F., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* 148, 1015–1028. doi: 10.1016/j.cell.2012.02.008

Hashimoto, K., Gross, B. G., and Lever, W. F. (1965). The ultrastructure of the skin of human embryos. I. The intraepidermal eccrine sweat duct. *J. Invest. Dermatol.* 45, 139–151. doi: 10.1038/jid.1965.110

Huang, J., Dattilo, L. K., Rajagopal, R., Liu, Y., Kaartinen, V., Mishina, Y., et al. (2009a). FGF-regulated BMP signaling is required for eyelid closure and to

specify conjunctival epithelial cell fate. *Development* 136, 1741–1750. doi: 10.1242/dev.034082

Huang, L., Pu, Y., Hu, W. Y., Birch, L., Luccio-Camelo, D., Yamaguchi, T., et al. (2009b). The role of Wnt5a in prostate gland development. *Dev. Biol.* 328, 188–199. doi: 10.1016/j.ydbio.2009.01.003

Kalaher, K. M., Anderson, W. I., and Scott, D. W. (1990). Neoplasms of the apocrine sweat glands in 44 dogs and 10 cats. *Vet. Rec.* 127, 400–403.

Kamberov, Y. G., Karlsson, E. K., Kamberova, G. L., Lieberman, D. E., Sabeti, P. C., Morgan, B. A., et al. (2015). A genetic basis of variation in eccrine sweat gland and hair follicle density. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9932–9937. doi: 10.1073/pnas.1511680112

Klaka, P., Grudl, S., Banowski, B., Giesen, M., Sattler, A., Proksch, P., et al. (2017). A novel organotypic 3D sweat gland model with physiological functionality. *PLoS One* 12:e0182752. doi: 10.1371/journal.pone.0182752

Kunisada, M., Cui, C. Y., Piao, Y., Ko, M. S., and Schlessinger, D. (2009). Requirement for Shh and Fox family genes at different stages in sweat gland development. *Hum. Mol. Genet.* 18, 1769–1778. doi: 10.1093/hmg/ddp089

Kurata, R., Futaki, S., Nakano, I., Fujita, F., Tanemura, A., Murota, H., et al. (2017). Three-dimensional cell shapes and arrangements in human sweat glands as revealed by whole-mount immunostaining. *PLoS One* 12:e0178709. doi: 10.1371/journal.pone.0178709

Leyden, J. J., Mcginley, K. J., Holzle, E., Labows, J. N., and Kligman, A. M. (1981). The microbiology of the human axilla and its relationship to axillary odor. *J. Invest. Dermatol.* 77, 413–416. doi: 10.1111/1523-1747.ep12494624

Liu, G., Liu, R., Li, Q., Tang, X., Yu, M., Li, X., et al. (2013). Identification of microRNAs in wool follicles during anagen, catagen, and telogen phases in Tibetan sheep. *PLoS One* 8:e77801. doi: 10.1371/journal.pone.0077801

Liu, N., Li, H., Liu, K., Yu, J., Cheng, M., De, W., et al. (2014). Differential expression of genes and proteins associated with wool follicle cycling. *Mol. Biol. Rep.* 41, 5343–5349. doi: 10.1007/s11033-014-3405-1

Liu, W., Wu, T., Dong, X., and Zeng, Y. A. (2017). Neuropilin-1 is upregulated by Wnt/beta-catenin signaling and is important for mammary stem cells. *Sci. Rep.* 7:10941. doi: 10.1038/s41598-017-11287-w

Lobitz, W. C. Jr., and Dobson, R. L. (1961). Dermatology: the eccrine sweat glands. *Annu. Rev. Med.* 12, 289–298. doi: 10.1146/annurev.me.12.020161.001445

Lu, C. P., Polak, L., Keyes, B. E., and Fuchs, E. (2016). Spatiotemporal antagonism in mesenchymal-epithelial signaling in sweat versus hair fate decision. *Science* 354:aah6102. doi: 10.1126/science.aah6102

Lv, X., Sun, W., Yin, J., Ni, R., Su, R., Wang, Q., et al. (2016). An Integrated Analysis of MicroRNA and mRNA expression profiles to identify RNA expression signatures in lambskin hair follicles in Hu sheep. *PLoS One* 11:e0157463. doi: 10.1371/journal.pone.0157463

Martin, A., Saathoff, M., Kuhn, F., Max, H., Terstegen, L., and Natsch, A. (2010). A functional ABCC11 allele is essential in the biochemical formation of human axillary odor. *J. Invest. Dermatol.* 130, 529–540. doi: 10.1038/jid.2009.254

Mayer, J. A., Foley, J., De La Cruz, D., Chuong, C. M., and Widelitz, R. (2008). Conversion of the nipple to hair-bearing epithelia by lowering bone morphogenetic protein pathway activity at the dermal-epidermal interface. *Am. J. Pathol.* 173, 1339–1348. doi: 10.2353/ajpath.2008.070920

Mellgren, A. M., Smith, C. L., Olsen, G. S., Eskiocak, B., Zhou, B., Kazi, M. N., et al. (2008). Platelet-derived growth factor receptor beta signaling is required for efficient epicardial cell migration and development of two distinct coronary vascular smooth muscle cell populations. *Circ. Res.* 103, 1393–1401. doi: 10.1161/CIRCRESAHA.108.176768

Moll, I., and Moll, R. (1992). Changes of expression of intermediate filament proteins during ontogenesis of eccrine sweat glands. *J. Invest. Dermatol.* 98, 777–785. doi: 10.1111/1523-1747.ep12499950

Morandi, F., Benazzi, C., and Simoni, P. (2005). Adenocarcinoma of apocrine sweat glands in a mouflon (*Ovis musimon*). *J. Vet. Diagn. Invest.* 17, 389–392. doi: 10.1177/104063870501700417

Narita, T., Saitoh, K., Kameda, T., Kuroiwa, A., Mizutani, M., Koike, C., et al. (2000). BMPs are necessary for stomach gland formation in the chicken embryo: a study using virally induced BMP-2 and Noggin expression. *Development* 127, 981–988.

Nie, Y., Li, S., Zheng, X., Chen, W., Li, X., Liu, Z., et al. (2018). Transcriptome reveals long non-coding RNAs and mRNAs involved in primary wool follicle induction in carpet sheep fetal skin. *Front. Physiol.* 9:446. doi: 10.3389/fphys.2018.00446

Nowak, J. A., Polak, L., Pasolli, H. A., and Fuchs, E. (2008). Hair follicle stem cells are specified and function in early skin morphogenesis. *Cell Stem Cell* 3, 33–43. doi: 10.1016/j.stem.2008.05.009

O'Shaughnessy, R. F., Christiano, A. M., and Jahoda, C. A. (2004). The role of BMP signalling in the control of ID3 expression in the hair follicle. *Exp. Dermatol.* 13, 621–629. doi: 10.1111/j.0906-6705.2004.00206.x

Owens, P., Bazzi, H., Engelking, E., Han, G., Christiano, A. M., and Wang, X. J. (2008). Smad4-dependent desmoglein-4 expression contributes to hair follicle integrity. *Dev. Biol.* 322, 156–166. doi: 10.1016/j.ydbio.2008.07.020

Plikus, M., Wang, W. P., Liu, J., Wang, X., Jiang, T. X., and Chuong, C. M. (2004). Morpho-regulation of ectodermal organs: integument pathology and phenotypic variations in K14-Noggin engineered mice through modulation of bone morphogenic protein pathway. *Am. J. Pathol.* 164, 1099–1114. doi: 10.1016/S0002-9440(10)63197-5

Preti, G., and Leyden, J. J. (2010). Genetic influences on human body odor: from genes to the axillae. *J. Invest. Dermatol.* 130, 344–346. doi: 10.1038/jid.2009.396

Purba, T. S., Haslam, I. S., Shahmalak, A., Bhogal, R. K., and Paus, R. (2015). Mapping the expression of epithelial hair follicle stem cell-related transcription factors LHX2 and SOX9 in the human hair follicle. *Exp. Dermatol.* 24, 462–467. doi: 10.1111/exd.12700

Rajkumar, V. S., Shiwen, X., Bostrom, M., Leoni, P., Muddle, J., Ivarsson, M., et al. (2006). Platelet-derived growth factor-beta receptor activation is essential for fibroblast and pericyte recruitment during cutaneous wound healing. *Am. J. Pathol.* 169, 2254–2265. doi: 10.2353/ajpath.2006.060196

Rezza, A., Sennett, R., Tanguy, M., Clavel, C., and Rendl, M. (2015). PDGF signalling in the dermis and in dermal condensates is dispensable for hair follicle induction and formation. *Exp. Dermatol.* 24, 468–470. doi: 10.1111/exd.12672

Roarty, K., and Serra, R. (2007). Wnt5a is required for proper mammary gland development and TGF-beta-mediated inhibition of ductal growth. *Development* 134, 3929–3939. doi: 10.1242/dev.008250

Roberts, D. J., Smith, D. M., Goff, D. J., and Tabin, C. J. (1998). Epithelial-mesenchymal signaling during the regionalization of the chick gut. *Development* 125, 2791–2801.

Rogers, G. E. (2006). Biology of the wool follicle: an excursion into a unique tissue interaction system waiting to be re-discovered. *Exp. Dermatol.* 15, 931–949. doi: 10.1111/j.1600-0625.2006.00512.x

Rompolas, P., and Greco, V. (2014). Stem cell dynamics in the hair follicle niche. *Semin. Cell Dev. Biol.* 2, 34–42. doi: 10.1016/j.semcdb.2013.12.005

Sato, K., Kang, W. H., Saga, K., and Sato, K. T. (1989). Biology of sweat glands and their disorders. I. Normal sweat gland function. *J. Am. Acad. Dermatol.* 20, 537–563. doi: 10.1016/S0190-9622(89)70063-3

Schlake, T. (2007). Determination of hair structure and shape. *Semin. Cell Dev. Biol.* 18, 267–273. doi: 10.1016/j.semcdb.2007.01.005

Sennett, R., Wang, Z., Rezza, A., Grisanti, L., Roitershtein, N., Sicchio, C., et al. (2015). An integrated transcriptome atlas of embryonic hair follicle progenitors, their niche, and the developing skin. *Dev. Cell* 34, 577–591. doi: 10.1016/j.devcel.2015.06.023

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shehadeh, N. H., and Kligman, A. M. (1963). The effect of topical antibacterial agents on the bacterial flora of the axilla. *J. Invest. Dermatol.* 40, 61–71. doi: 10.1038/jid.1963.10

Sun, T. T., Shih, C., and Green, H. (1979). Keratin cytoskeletons in epithelial cells of internal organs. *Proc. Natl. Acad. Sci. U.S.A.* 76, 2813–2817. doi: 10.1073/pnas.76.6.2813

ten Dijke, P., and Hill, C. S. (2004). New insights into TGF-beta-Smad signalling. *Trends Biochem. Sci.* 29, 265–273. doi: 10.1016/j.tibs.2004.03.008

Toyoda, Y., Sakurai, A., Mitani, Y., Nakashima, M., Yoshiura, K., Nakagawa, H., et al. (2009). Earwax, osmidrosis, and breast cancer: why does one SNP (538G > A) in the human ABC transporter ABCC11 gene determine earwax type? *FASEB J.* 23, 2001–2013. doi: 10.1096/fj.09-129098

Toyoda, Y., Takada, T., Gomi, T., Nakagawa, H., Ishikawa, T., and Suzuki, H. (2017). Clinical and molecular evidence of ABCC11 protein expression in axillary apocrine glands of patients with axillary osmidrosis. *Int. J. Mol. Sci.* 18:E417. doi: 10.3390/ijms18020417

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Wegner, J., Loser, K., Apsite, G., Nischt, R., Eckes, B., Krieg, T., et al. (2016). Laminin alpha5 in the keratinocyte basement membrane is required for epidermal-dermal intercommunication. *Matrix Biol.* 56, 24–41. doi: 10.1016/j.matbio.2016.05.001

Wilke, K., Martin, A., Terstegen, L., and Biel, S. S. (2007). A short history of sweat gland biology. *Int. J. Cosmet. Sci.* 29, 169–179. doi: 10.1111/j.1467-2494.2007.00387.x

Xu, M., Horrell, J., Snitow, M., Cui, J., Gochnauer, H., Syrett, C. M., et al. (2017). WNT10A mutation causes ectodermal dysplasia by impairing progenitor cell proliferation and KLF4-mediated differentiation. *Nat. Commun.* 8:15397. doi: 10.1038/ncomms15397

Yanagawa, S., Yokozeki, H., and Sato, K. (1986). Origin of periodic acid-Schiff-reactive glycoprotein in human eccrine sweat. *J. Appl. Physiol.* 60, 1615–1622. doi: 10.1152/jappl.1986.60.5.1615

Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., et al. (2006). A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat. Genet.* 38, 324–330. doi: 10.1038/ng1733

Yue, Y., Guo, T., Yuan, C., Liu, J., Guo, J., Feng, R., et al. (2016). Integrated analysis of the roles of long noncoding RNA and coding RNA expression in sheep (*Ovis aries*) skin during initiation of secondary hair follicle. *PLoS One* 11:e0156890. doi: 10.1371/journal.pone.0156890

Zouvelou, V., Luder, H. U., Mitsiadis, T. A., and Graf, D. (2009). Deletion of BMP7 affects the development of bones, teeth, and other ectodermal appendages of the orofacial complex. *J. Exp. Zool. B Mol. Dev. Evol.* 312B, 361–374. doi: 10.1002/jez.b.21262

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Zheng, Nie, Chen, Liu, Tao, Hu, Hu, Qiao, Qi, Pei, Cai, Yu and Mou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 March 2019
doi: 10.3389/fgene.2019.00145

Genetic Basis of Phenotypic Differences Between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-Specific Expression in Their F1 Hybrids

Yanhong Cao[1,2†], Han Xu[1†], Ran Li[1†], Shan Gao[1], Ningbo Chen[1], Jun Luo[1] and Yu Jiang[1*]*

[1] Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, [2] Guangxi Key Laboratory of Livestock Genetic Improvement, The Animal Husbandry Research Institute of Guangxi Zhuang Autonomous Region, Nanning, China

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:
George E. Liu,
Agricultural Research Service (USDA),
United States
Shikai Liu,
Ocean University of China, China

***Correspondence:**
Yu Jiang
yu.jiang@nwafu.edu.cn
Jun Luo
luojun@nwafu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 25 August 2018
Accepted: 12 February 2019
Published: 05 March 2019

Citation:
Cao Y, Xu H, Li R, Gao S, Chen N, Luo J and Jiang Y (2019) Genetic Basis of Phenotypic Differences Between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-Specific Expression in Their F1 Hybrids. Front. Genet. 10:145. doi: 10.3389/fgene.2019.00145

Chinese Yunling black goats and African Nubian goats are divergent breeds showing significant differences in body size, milk production, and environmental adaptation. However, the genetic mechanisms underlying these phenotypic differences remain to be elucidated. In this report, we provide a detailed portrait of allele-specific expression (ASE) from 54 RNA-Seq analyses across six tissues from nine F1 hybrid offspring generated by crossing the two breeds combined with 13 genomes of the two breeds. We identified a total of 524 genes with ASE, which are involved in bone development, muscle cell differentiation, and the regulation of lipid metabolic processes. We further found that 38 genes with ASE were also under directional selection by comparing 13 genomes of the two breeds; these 38 genes play important roles in metabolism, immune responses, and the adaptation to hot and humid environments. In conclusion, our study shows that the exploration of genes with ASE in F1 hybrids provides an efficient way to understand the genetic basis underlying the phenotypic differences of two diverse goat breeds.

Keywords: allele-specific expression (ASE), Chinese Yunling black goat, Nubian goat, whole genome sequencing, RNA-seq

INTRODUCTION

A domestic species can present diverse phenotypic differences due to the adaptation to local environments and artificial selection. Yet, it has been difficult to identify the causative genes that contribute to these phenotypic differences. Some studies have relied on genomic selection signals (Dong et al., 2012; Benjelloun et al., 2015). However, the identified selection signals generally contain a high proportion of background noise. Comparative transcriptome analysis of breeds with distinct traits is another frequently used approach (Hayano-Kanashiro et al., 2009; von Heckel et al., 2016). However, the resulting differentially expressed genes reflect both cis-acting and trans-acting regulatory variations, thus presenting little power to characterize the genetic architecture and identify causative genes. With the development of sequencing-based methods to study the transcriptome, it is possible to make use of natural sequence variation to trace and quantify allele-specific expression (ASE) in F1 hybrid individuals generated from crosses of two different lines of interest (Crowley et al., 2015; Aguilar-Rangel et al., 2017). Characterization of ASE in F1

material avoids the problem of comparing parents that may differ dramatically in their growth and development by evaluating both alleles within the same cellular environment, directly revealing cis-acting genetic variation related to transcript accumulation.

The black goats in Southwest China are characterized by a tolerance to crude feed, a higher resistance to parasitic diseases, and thinner muscle fibers (Miao et al., 2015). However, the growth rate of these black goats is much slower than that of commercialized breeds (Zhao et al., 2011) improved by European countries. Nubian goats, a popular commercialized breed, exhibit high feed efficiency and a fast growth rate but are susceptible to parasites (Kholif et al., 2017; Rahmatalla et al., 2017), which are common in the hot and humid environment of South China. In the past decades, Nubian goats have been continuously imported into China to improve the production performance of local breeds (Yuan et al., 2017). Understanding the genetic basis underlying the distinct phenotypes of these two breeds will be a perquisite for new breed selection and customizing strategies for cross breeding.

To understand the genetic mechanisms underlying the phenotypic differences between these two breeds, an F1 hybrid population was generated by crossing female Chinese Yunling black goats and male Nubian goats, and the transcriptomes of the F1 hybrids were analyzed in six tissues (liver, bone, muscle, fat, skin, and mammary gland tissues) to detect ASE. Combined with the selection signals identified in Chinese Yunling black goats and Nubian goats, we provide further insights into the genomic contributions underlying the phenotypic diversity between these two goat breeds.

MATERIALS AND METHODS

Sample Collection

Six Chinese Yunling black ewes and four Nubian rams were selected to produce nine F1 hybrids. Nine female F1 hybrids (three from each cross) were slaughtered after being stunned by high voltage electricity. Liver, bone, muscle, fat, skin, and mammary gland tissues from all nine hybrids were rapidly dissected, snap frozen in liquid nitrogen, and stored at −80°C until use. In addition, three of the individuals additionally collected horns, hooves, and rumen for the calculation of reads counts and genotype judgment. For each tissue sample, two replicates were collected simultaneously. Blood samples were collected from the parents (six Chinese Yunling black ewes and four Nubian rams) and three female F1 hybrids.

DNA Extraction and DNA Sequencing

Genomic DNA was extracted from blood samples using a Tiangen DNA isolation kit (Tiangen Biotech, Beijing, China). At least 6 μg of genomic DNA from each sample was used to construct a sequencing library following Illumina instructions. Paired-end sequencing libraries with an insert size of approximately 500 bp were sequenced using an Illumina HiSeq 2000 (Berry Genomics Company).

RNA Extraction and RNA-Sequencing

Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer protocols. RNA quality was measured using an Agilent 2100 Bioanalyzer. All samples had an RNA integrity Number (RIN) ≥7. Library construction and sequencing were performed according to Illumina instructions. mRNA was isolated from DNA-free total RNA using the Dynabeads mRNA DIRECT Kit (Invitrogen) and fragmented. First-strand cDNA was generated using Random Primer p(dN)6 and Superscript III, after which second-strand cDNA synthesis and adaptor ligation were performed. cDNA fragments of 400–500 bp were isolated. The library was sequenced using the Illumina X-ten platform to generate 150 bp paired-end reads (Berry Genomics Company).

Genomic Sequence Analysis

Before alignment, the raw data were processed to filter out adaptors and low-quality reads. High-quality clean reads from the DNA sequencing of parents were aligned to the goat reference genome (Bickhart et al., 2017) using BWA software (Li and Durbin, 2009). We then assigned SNPs to the two groups using the Genome Analysis Toolkit (GATK, v3.2-2) (McKenna et al., 2010) to discriminate the parents from both lines. Next, we filtered low-quality sites using the parameter QUAL <30. All the assigned variants were annotated using the package ANNOVAR (Version: 2013-08-23) (Wang et al., 2010).

Analysis of Selective Sweeps

We performed a selective sweep analysis by calculating the genetic differentiation (Fst) and heterozygosity (Hp) of each 150 KB genome window and 75 KB step length. Fst was calculated using VCFtools (Kofler et al., 2011), and Hp was calculated as described previously (Rubin et al., 2012). The Hp and Fst values were converted to a standard normal distribution, denoted by ZHp and ZFst. In addition, regions that exhibited low Hp and high Fst values were screened as candidates. To understand the biological functions of genes within candidate regions, GO analysis was performed.

Transcriptome Mapping and Quantification of Expression

Clean reads were mapped to the CHIR_3.0 reference genome (Bickhart et al., 2017) using STAR with default options. Next, the unmapped reads were remapped to the genome using Hisat2 (v 2.0.3) (Pertea et al., 2016). The assignment of reads to genes was performed using StringTie (Pertea et al., 2016). The expression levels of the protein-coding genes were quantified using the R package "Ballgown" (Pertea et al., 2016).

ASE Analysis

Allele counts were retrieved using a homemade Python scripts (**Supplementary File 1**: GetSnpCountFromBam.py) which calculates allele counts at SNP positions. Heterozygous sites with individual allele read depth <20 and total (both alleles) read depth <50 were filtered out. A binomial test and Benjamini-Hochberg FDR correction were performed. Cut-off criteria of allele ratio >0.7 or <0.3 and FDR < 0.05 were used to

identify significant allelic imbalances. Previously identified imprinted genes obtained from an online database (http://www.geneimprint.com/site/home) were excluded from our final gene set.

Gene Ontology Analysis

GO pathway enrichment analyses were performed to identify enriched functions in KOBAS3.0 (http://kobas.cbi.pku.edu.cn/index.php). We converted the goat gene symbol IDs into human homologous gene symbol IDs using Blastp before performing GO pathway analyses, as the goat gene annotations in the KOBAS3.0 database were inadequate. We set the EASE value to 0.05 for the enrichment analysis.

RESULTS

Genomic Variants of Chinese Yunling Black Goats and Nubian Goats

Six Chinese Yunling black goats, four Nubian goats and three F1 hybrids were selected for genome resequencing (**Figure 1A**). The genome resequencing achieved an average depth of 15X and a mapping rate of 99.54%. A total of 11.52 million SNPs were found to differ between the parents of each breed, and 309,984 SNPs were expressed. Among the total discriminating SNPs, 0.85% were detected in coding regions (**Figure 1B**; **Supplementary File 2**). In addition, we detected 313 SNPs in termination codons and 258 SNPs in splice sites. We found that 24,701 genes were annotated genes with at least one discriminating SNP.

Interestingly, we found 7,365 genes that contained more than 100 SNPs, implying that these genes were highly diverse and might be particularly susceptible to artificial selection (**Figure 1**, **Supplementary Figure 1A**). The proportion of substitution transitions (69.4%) was much higher than that of transversions (30.7%) (**Supplementary Figure 1B**). The transition:transversion ratio was 2.26:1, which is similar to that found in other goat studies (Guan et al., 2016).

Genome-Wide Selective Sweep Analysis

F-statistic (Fst) scores were calculated to measure the signature of selection between the Yunling black goats and Nubian goats. We scanned the autosomes with a nonoverlapping 100 kb window and calculated the Fst value for each window. We focused on the regions with extremely high Z-transformed Fst values (Top %1) in the genome-wide empirical distribution. In total, 250 putative selective sweep regions containing 521 candidate genes were identified (**Figure 2A**; **Supplementary File 3**).

The region with the strongest differentiation signal [ZFst = 7.85] between the two breeds was the 13.73–13.95 MB region of chromosome 12, which contained *LOC108637252/LOC108637248/LOC102180841/LOC102180583* (*MRP4*). The product of this region protects cells against toxicity by acting as an ion efflux pump, in addition to influencing dendritic cell migration (Li et al., 2017). We also identified several genes showing differentiation, including

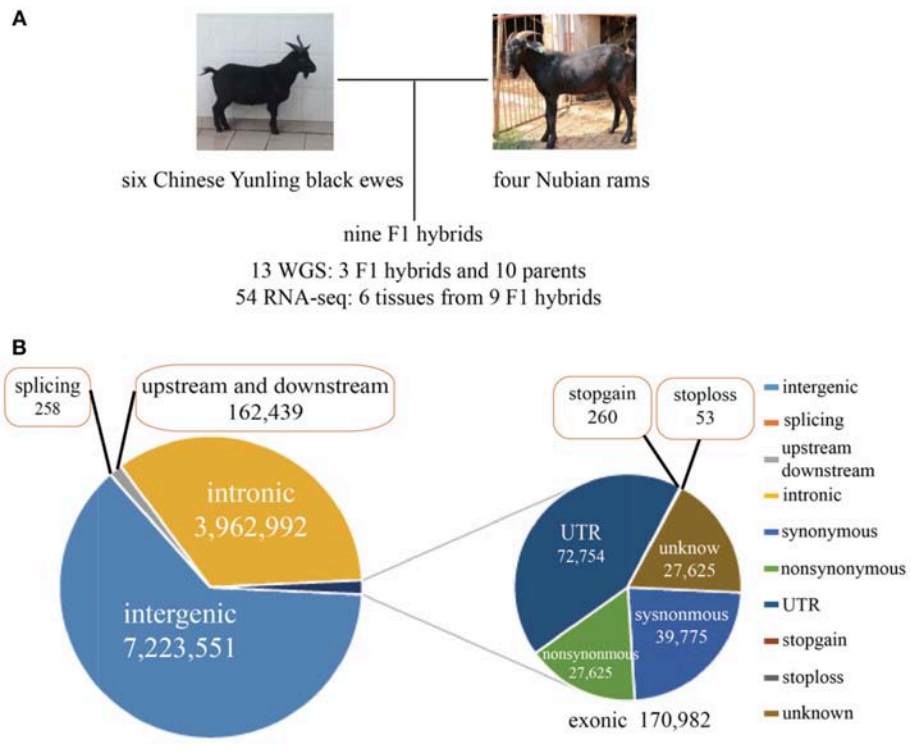


FIGURE 1 | (A) Experimental design and brief data summary. **(B)** Statistics regarding the location and mutation type of the whole-genome SNPs.

CELF2, *TDO2*, *ZFPM1*, *TAP1*, *LOC102177333* (*CYP2D6*), and *LOC102173339* (*CYP8B1*).

Heterozygosity (Hp) was also used to detect putative selective sweeps. The distribution of the observed Hp values and the Z transformations of Hp and ZHp are plotted in **Figure 2**. We searched the regions with the lowest heterozygosity (top 1% based on |ZHp| scores), which yielded a total of 275 putative selective sweep regions containing 785 candidate genes in Nubian goats (**Figure 2B**) and 270 putative selective sweep regions containing 854 candidate genes in Chinese Yunling black goats (**Figure 2C** and **Supplementary File 3**).

We observed high ZHp values (ZHp = 6.40) across the *PCDHB* (protocadherin B) gene family (**Figure 2B**) in Nubian goats and for the *UBR4* and *EMC1* genes (ZHp = 6.28) in Chinese Yunling black goats (**Figure 2C**).

Transcriptome Characterization of F1 Hybrids

To detect genes with ASE and infer the existence of cis-regulatory variants, we combined the RNA-Seq data from six tissues (liver, bone, muscle, fat, skin, and mammary gland tissues) and the whole-genome sequencing results from three female F1 hybrids (**Figure 1A** and **Supplementary File 2**). The whole-genome resequencing data were used to exclude possible base changes in RNA sequences resulting from RNA editing.

The greatest number of ASE SNPs (2,685) were detected in the mammary gland, while 1,556 were detected in muscle **Table 1**. Most of the ASE SNPs were located in annotated genes (liver 79.6%, bone 81.7%, muscle 87.5%, fat 77.8%, mammary gland 80.9%, and skin 81.4%) (**Table 1**). The percentage of synonymous regions was over 50% in the six tissues, and the ratio in muscle was highest, reaching 64%(**Table 2**).

The genes we identified that had ASE may have included imprinted genes. We therefore collected the imprinted genes of human, mouse, cattle and sheep (**Supplementary Table 1**) from a publicly available database (http://www.geneimprint.com/site/home) and excluded these from our results. In this way, we finally identified 524 genes with ASE in the six tissues, ranging from 78 in muscle to 144 in liver (**Table 3**). The greatest number of ASE genes comprised protein-coding genes, whose proportion in the six tissues of all hybrids was above 90%, followed by noncoding RNAs (**Table 3**). Using an FPKM > 0.01 as a threshold, the average expression

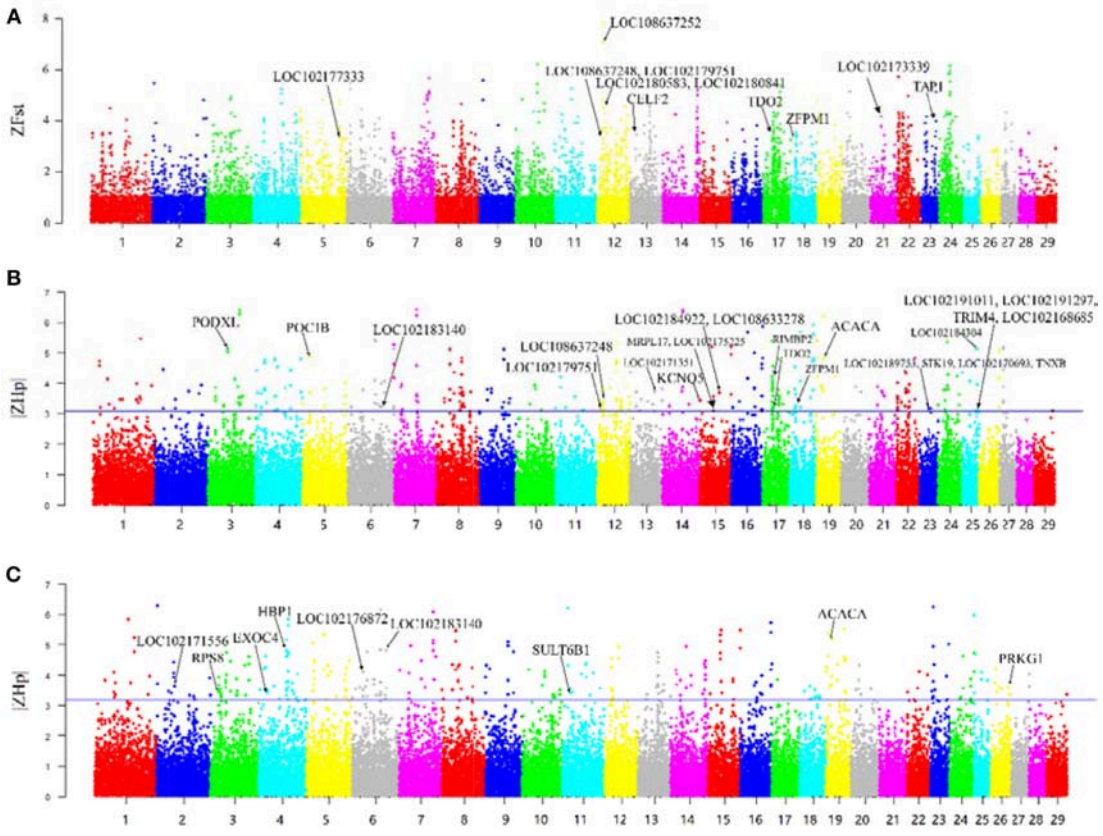


FIGURE 2 | Overview of selective sweeps in the Nubian and Yunling black goat breeds based on ZFst and ZHp values. The labeled genes in bold characters represent those genes with selective signals that overlap with ASE genes. **(A)** ZFst values between Nubian and Yunling black goat breeds. Bold names are the ASE genes contained in the maximum Z-Fst in the 100 kb window. **(B)** ZHp value of Nubian goats. Bold names are the ASE genes contained in the minimum |ZHp| group of Nubian goats in the 100 kb window. **(C)** ZHp value of Yunling Black goats. Bold names are the ASE genes contained in the minimum |ZHp| group of Yunling Black goats in the 100 kb window.

TABLE 1 | Annotation of SNPs with ASE from six tissues.

ASE SNP annotation	Liver	Bone	Muscle	Fat	Mammary	Skin
Total	2,341	2,004	1,556	2,349	2,685	2,362
SNPs in annotated genes	1,864	1,637	1,361	1,827	2,173	1,923
Exon	885	1,073	1,103	1,347	1,353	1,493
Intron	978	562	257	477	817	429
Splice region	1	2	1	3	3	1
Intergenic	381	257	146	406	400	325
Upstream and downstream	96	110	49	116	112	114

TABLE 2 | Mutation statistics of SNPs with ASE located on exons.

	Liver	Bone	Muscle	Fat	Mammary	Skin
Synonymous	230	291	323	362	384	397
Non-synonymous	189	266	182	244	277	244
Stopgain	0	0	0	0	0	0
Stoploss	2	0	0	0	0	0

TABLE 3 | Encoding type for genes with ASE.

	Protein coding	Noncoding RNA	Pseudogene	Total
Liver	135	8	1	144
Bone	101	3	3	107
Muscle	74	4	0	78
Fat	92	2	3	97
Mammary	114	7	3	124
Skin	121	4	2	127

level of the ASE gene was significantly higher than the average expression of the normal gene in the same tissue (**Supplementary Table 2**).

Functional Annotation of ASE Genes

To explore the tissue specificity of the genes with ASE, functional enrichment analyses were performed. In bone, 276 Gene Ontology (GO) terms were significantly enriched in 77 ASE genes ($P < 0.05$), most of which were associated with hematopoietic or lymphoid organ development (**Table 4** and **Supplementary File 3**). There were 332GO terms enriched in 72 ASE genes in muscle ($P < 0.05$) (**Supplementary File 4**), which were mainly involved in striated muscle cell development, the actin cytoskeleton, muscle cell differentiation, actin-mediated cell contraction, muscle fiber development, striated muscle thin contraction, and muscle tissue development (**Table 4**). The results revealed that 385 GO terms were significantly enriched in 82 ASE genes in fat tissue ($P < 0.05$) (**Supplementary File 4**). Interestingly, lipid-related processes (lipid localization, lipid binding, and lipid metabolic processes) were significantly

TABLE 4 | GO analysis of the genes with ASE in difference tissues.

Tissue	GO Term	Gene number	Corrected *P*-value
Bone	Hematopoietic or lymphoid organ development	8	1.88E-03
	Hemopoiesis	7	5.91E-03
Muscle	Striated muscle cell development	10	1.77E-11
	Actin cytoskeleton	13	4.02E-10
	Muscle cell differentiation	11	5.19E-09
	Actin-mediated cell contraction	6	1.88E-06
	Muscle fiber development	5	3.37E-06
	Striated muscle contraction	6	1.68E-05
	Muscle tissue development	6	7.06E-04
Fat	Lipid localization	5	6.23E-03
	Lipid binding	6	1.30E-02
	Lipid metabolic process	8	2.84E-02
Skin	Keratin filament	11	2.32E-13
Liver	Small molecule metabolic process	43	1.39E-23
	Organic acid metabolic process	31	4.47E-21
	Cellular metabolic process	78	1.18E-18
	Carboxylic acid metabolic process	26	7.63E-17
	Oxoacid metabolic process	26	8.50E-17
	Monocarboxylic acid metabolic process	21	2.16E-15
	Cofactor metabolic process	17	1.63E-13
	Organonitrogen compound metabolic process	33	2.65E-13
	Cellular lipid metabolic process	22	1.74E-11
	Coenzyme metabolic process	14	3.52E-11

enriched in 11 ASE genes, including ABCA6, LBP, SEC14L2, SERINC2, ACACA, CD36, AADAC, C3, LGALS12, PLA2G16, and MFGE8 (**Table 4**). The liver is an important metabolic organ, for which 10 metabolic-related GO terms were found, including the metabolic processes of small molecules, organic acids, cellular functions, carboxylic acid, oxoacid, monocarboxylic acid, cofactors, organonitrogen compounds, cellular lipids, and coenzyme function (**Table 4**; **Supplementary File 4**). There was only one significantly enriched GO term for bone and skin, respectively, and none was significantly enriched for mammary gland.

Among the six tissues, the ASE genes of the liver presented the highest tissue specificity (77.8%), followed by those in the skin (71.7%), bone (68.2%), muscle (67.9%), and mammary tissue (52.4%). In contrast, fat exhibited a lower level of tissue specificity (45.4%). **Supplementary Table 3** shows the status of the overlap of ASE genes in different tissues. For example, *HLA-A*, *HLA-B*, *HLA-DQA1*, *HLA-DQB1*, and *LOC106503915* were detected in all tissues **Supplementary Table 3**. The *HLA* genes encoded major histocompatibility complex (*MHC*) class I proteins in the context of specific cell surfaces (Valenzuela-Ponce et al., 2018), involving *HLA-A*, *HLA-B*, *HLADRB1*, *HLA-DQA1*, and *HLA-DQB1* (Emerson et al., 2017). *HLA* played a major role in the control of the immune response and its

associations with a wide variety of immunological and infectious disorders, such as type I diabetes, multiple sclerosis, rheumatoid arthritis, Grave's disease, ankylosing spondylitis, and systemic lupus erythematosus (Spínola et al., 2016).

Potential Effects of ASE Genes Under Selection

We also observed several ASE genes under selection. The range of selected ASE genes in six tissues was 5 to 13. Among these genes, we highlight the Ribosomal protein S8 (*RPS8*) gene. *RPS8* was detected in four tissues: skin, bone, mammary tissues and muscle tissue. It also showed the lowest level of heterozygosity in the selected region in Yunling black goats (|ZHp| = 3.46). *RPS8* has been used to develop a species-specific PCR-RFLP diagnostic tool for ovine babesiosis and theileriosis, which are hemoprotozoal diseases that cause economic losses among sheep and goats in tropical and subtropical regions (Tian et al., 2013).

Another ASE gene, Multidrug resistance protein 4 (*MRP4*), was specifically expressed in the bone, which contained 32 SNPs (29 exonic and three intronic SNPs). Based on the results for *MRP4*, Nubian goats with the highest Fst value [Z(Fst) = 7.08] and Hp value (|ZHp| = 3.33) were selected. *MRP4* has been identified as an important transporter for signaling molecules, including cyclic nucleotides and several lipid mediators in platelets. *MRP4* is known to play a critical role in the elimination of numerous drugs, carcinogens, toxicants, and their conjugated metabolites and is expressed at the basolateral surface of hepatocytes, which can facilitate cellular efflux to sinusoidal blood for entry into the systemic circulation (Li et al., 2017).

We further examined the ASE genes under selection. Thirty-eight genes with ASE from the six tissues were identified as being under directional selection, implying that they are involved in biologically essential functions, and these genes were therefore defined as core-ASE genes. The largest number of core-ASE genes (13) was found in the liver, where four genes were associated with metabolism (*LOC102191011*, *LOC102177333*, *LOC102173339*, and *LOC102191297*), and another four were associated with the immune system (*TAP1*, *LOC102189753*, *LOC102171351*, and *TDO2*). In skin, there were four core-ASE genes, including *TNXB*, which is related to the pathogenesis of systemic lupus erythematosus and *RPS8*, which is involved in resistance to hemoprotozoal diseases that cause economic losses among sheep and goats in tropical and subtropical regions. A small number of core-ASE genes were also found in bone (6) and the mammary glands (3), as shown in (**Figure 3**). Notably, two core-ASE genes (*PODXL* and *RPS8*) were present in at least two tissues, and in the *PODXL* gene, the ASE phenomenon was detected in eight SNPs as soon as the SNPs became heterozygotic in individuals, except for SNP1 and SNP2 in sample 1 (**Figure 4**).

DISCUSSION

Studies in recent years have shown that ASE analysis is an efficient tool for identifying causative genetic variations. However, few studies have conducted ASE analysis in livestock, partially due to the high costs of generating hybrids. In this study, we generated F1 hybrids of two diverse breeds and then explored ASE genes in these hybrids. We further combined genomic selection signals and ASE analysis to gain insight into the genomic contributions underlying phenotypic differences and local adaptability to different environments. It should be noted that although we did not have reciprocal cross for the identification of imprinted genes, we excluded the candidates based on previously identified imprinted genes as much as possible in other species (human, mice, cattle and sheep).

Genetic diversity patterns and overall low heterozygosity are commonly used statistical methods for detecting genomic regions related to selection in domesticated animals. To detect putative selective loci in the present study, we performed

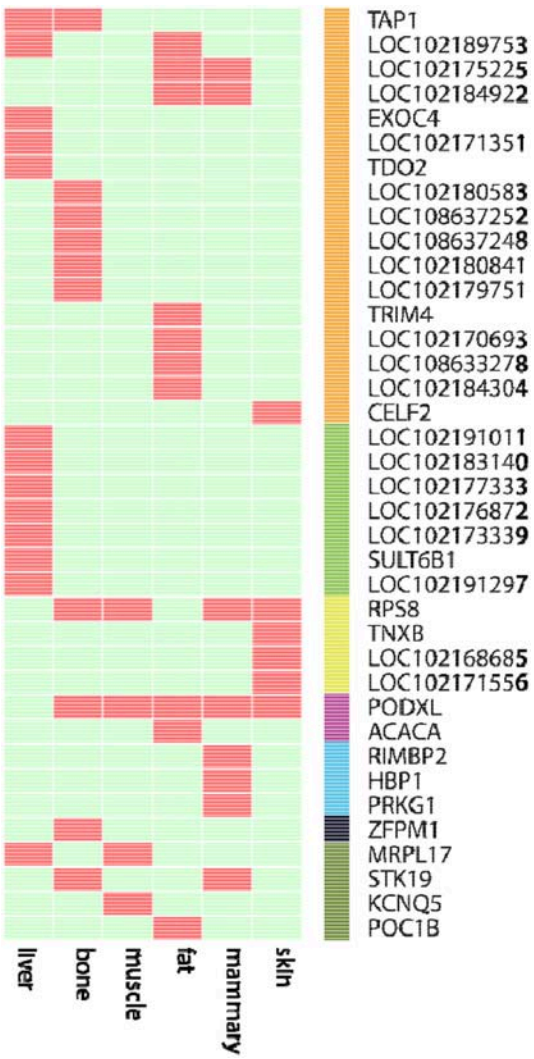


FIGURE 3 | The 38 core-ASE genes and their functions. Red and light green indicate the presence and absence, respectively, of genes with ASE (rows) in the corresponding tissues (columns). The right bar indicates the biological functions of the corresponding genes (Orange: immune responses; green: metabolism; yellow: adaptation to hot and humid environments; purple: functions associated with body measurements and weight; cyan: cell regulation metabolism; black: hematopoiesis; dark green: undefined).

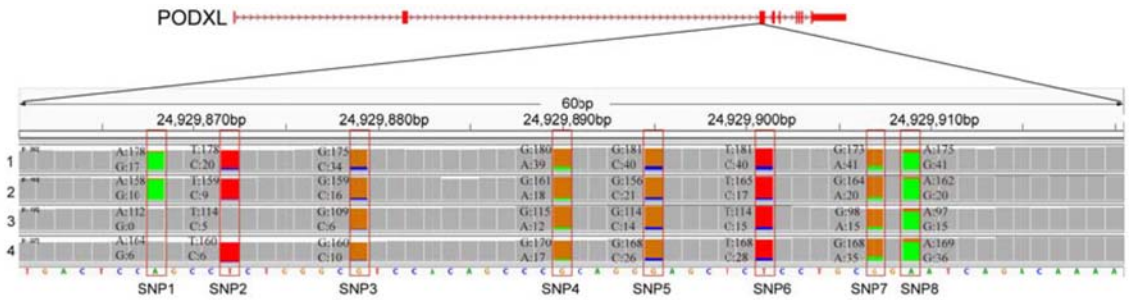


FIGURE 4 | The allelic expression of eight SNPs containing ASE sites within the third exon of PODXL gene. The allele count information for each base was shown for the muscle tissues from four individuals (tracks 1–4). The colored bar indicates ASE sites with different colors representing the counts of corresponding alleles.

sequencing in six Chinese Yunling black goats (representative of Chinese southern domestic black goats) and four Nubian goats and calculated the corresponding Fst and Hp values. Top 1% of the selection signals does not automatically mean positive selection, but they could narrow down our list of candidate genes. Thus, we identified a total of 521 genes showing population differentiation that potentially contribute to the phenotypic and adaptation traits of the goats. However, the genetic differentiation between the two breeds may be due to breeding, evolutionary and management history. Furthermore, the adaptation and phenotypic differences of the goats may be mediated by a complex network of genes that act in tandem, rather than by the action of single candidate gene (Lv et al., 2014; Kim et al., 2015). It is therefore difficult to directly draw conclusions regarding the genetic mechanisms underlying the observed traits based only on genomic selection signals. With only 3 F1, 6 Yunling and 4 Nubian WGS, it looks sample size ($n = 13$) was too small to obtain reliable estimates. But our objective is to understand the phenotypic difference among Yunling and Nubian breed by tracing and quantifying allele-specific expression (ASE) in F1 hybrid individuals. We are more concerned about the ASE results since this part is based on RNA-seq data from multiple tissues. The selection signal analysis indeed could have a large number of false positive results. However, our reported selection signals were only used to narrow down the candidate genes from ASE analysis by choosing the overlapping genes. Therefore, although the selection signals would contain large proportion of false positives, it would not affect our interpretation of the main results.

ASE is an important source of phenotypic diversity. The phenotypic traits of F1 hybrids are determined by the coordinated expression of alleles from both parents. In F1 hybrids, the two alleles from the parents will be exposed to the same trans-acting factors; thus, allelic-specific expression can be attributed only to differences in the cis-acting factors. When the parents show high genetic divergence, we can easily distinguish the origin of the two alleles based on their inherited SNPs from their parents. In the characterization of ASE in F1 hybrids, both alleles within the same cellular environment are evaluated, directly revealing cis-acting genetic variation in transcript accumulation (Springer and Stupar, 2007; Perumbakkam et al., 2013; Aguilar-Rangel et al., 2017). To our knowledge, this is the first ASE analysis of an F1 hybrid generated from a cross between two different breeds of goats.

We set up an experimental cross designed to characterize the divergence of gene regulation between Chinese Yunling black and Nubian goats via RNA-Seq analysis in six tissues of nine F1 hybrids, to identify candidate genes underlying local adaptation. Hundreds of ASE genes were found in different tissues, and a small proportion of these genes (core-ASE) were further shown to experience directional selection. These core-ASE genes are related to many essential biological processes, including metabolism (*LOC102191011*, *LOC102177333*, *LOC102173339*, and *LOC102191297*) (Elens et al., 2012; Bertaggia et al., 2017; Buermans Henk et al., 2017), immune responses (*TAP1*, *LOC102189753*, *LOC102171351,* and *TDO2*) (Grassmann et al., 2016; Hanalioglu et al., 2017; Kota et al., 2017), and the adaptation to hot and humid environments (RPS8 and TNXB) (Wei and Hemmings, 2003). The identification of these genes will help to explain the phenotypic differences and genetic mechanisms underlying the adaptation of the two representative goat breeds examined in this study and will supply a theoretical basis for crossbreeding and the improvement, breeding and the selection of local goats.

DATA AVAILABILITY

The whole genome sequence data and RNA-seq data have been deposited in NCBI short read archive under study PRJNA504493 and PRJNA485657, respectively.

ETHICS STATEMENT

All animal experiments were approved by the Institutional Animal Care and Use Committee at the College of Animal Science and Technology, Northwest A & F University. All experimental goats were housed in Black Goat Farm of Guangxi Institute of Animal Science, Nanning, China.

AUTHOR CONTRIBUTIONS

YJ, YC, and JL designed the experiment. YC fed the goats and collected the experimental tissues. HX and RL contributed to analyzing the data and interpreting the results. YC wrote

the manuscript with input from all the authors. SG and NC participated in designing the structure of the article. All the authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (31572381), the National Thousand Youth Talents Plan, Guangxi Animal Husbandry Technology Project (201633013), Guangxi Science and Technology Major Project (AA18118041) and Guangxi Cattle and Goat Innovation Team (nycytxgxcxtd-09-02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00145/full#supplementary-material

REFERENCES

Aguilar-Rangel, M. R., Chávez Montes, R. A., González-Segovia, E., Ross-Ibarra, J., Simpson, J. K., and Sawers, R. J. H. (2017). Allele specific expression analysis identifies regulatory variation associated with stress-related genes in the Mexican highland maize landrace Palomero Toluqueño. *PeerJ.* 5:e3737. doi: 10.7717/peerj.3737

Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F., Coissac, E., Stucki, S., et al. (2015). Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Front. Genet.* 6:107. doi: 10.3389/fgene.2015.00107

Bertaggia, E., Jensen, K. K., Castro-Perez, J., Xu, Y., Di Paolo, G., Chan, R. B., et al. (2017). Cyp8b1 ablation prevents Western diet-induced weight gain and hepatic steatosis because of impaired fat absorption. *Am. J. Physiol. Endocrinol. Metab.* 313, E121–E133. doi: 10.1152/ajpendo.00409.2016

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi: 10.1038/ng.3802

Buermans Henk, P. J., Vossen Rolf, H. A. M., Anvar Seyed, Y., Allard William, G., Guchelaar, H. J., White Stefan, J., et al. (2017). Flexible and scalable full-length CYP2D6 long amplicon PacBio sequencing. *Hum. Mutat.* 38, 310–316. doi: 10.1002/humu.23166

Crowley, J. J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I. K., Kim, Y., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* 47, 353–360. doi: 10.1038/ng.3222

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2012). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31, 135–141. doi: 10.1038/nbt.2478

Elens, L., van Gelder, T., Hesselink, D. A., Haufroid, V., and van Schaik, R. H. N. (2012). CYP3A4*22: promising newly identified CYP3A4 variant allele for personalizing pharmacotherapy. *Pharmacogenomics* 14, 47–62. doi: 10.2217/pgs.12.187

Emerson, R. O., DeWitt, W. S., Vignali, M., Gravley, J., Hu, J. K., Osborne, E. J., et al. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* 49, 659–665. doi: 10.1038/ng.3822

Grassmann, F., Cantsilieris, S., Schulz-Kuhnt, A.-S., White, S. J., Richardson, A. J., Hewitt, A. W., et al. (2016). Multiallelic copy number variation in the complement component 4A (C4A) gene is associated with late-stage age-related macular degeneration (AMD). *J. Neuroinflamm.* 13:81. doi: 10.1186/s12974-016-0548-0

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, Y., Na, R., et al. (2016). Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). *Sci. Rep.* 6:36372. doi: 10.1038/srep36372

Hanalioglu, D., Ayvaz, D. C., Ozgur, T. T., van der Burg, M., Sanal, O., and Tezcan, I. (2017). A novel mutation in TAP1 gene leading to MHC class I deficiency: report of two cases and review of the literature. *Clin. Immunol.* 178, 74–78. doi: 10.1016/j.clim.2017.01.011

Hayano-Kanashiro, C., Calderón-Vázquez, C., Ibarra-Laclette, E., Herrera-Estrella, L., and Simpson, J. (2009). Analysis of gene expression and physiological responses in three mexican maize landraces under drought stress and recovery irrigation. *PLoS ONE* 4:e7531. doi: 10.1371/journal.pone.0007531

Kholif, A. E., Abdo, M. M., Anele, U. Y., El-Sayed, M. M., and Morsy, T. A. (2017). *Saccharomyces cerevisiae* does not work synergistically with exogenous enzymes to enhance feed utilization, ruminal fermentation and lactational performance of Nubian goats. *Livest. Sci.* 206, 17–23. doi: 10.1016/j.livsci.2017.10.002

Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., et al. (2015). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255–264. doi: 10.1038/hdy.2015.94

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., et al. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* 6:e15925. doi: 10.1371/journal.pone.0015925

Kota, S. K., Pernicone, E., Leaf, D. E., Stillman, I. E., Waikar, S. S., and Kota, S. B. (2017). BPI fold-containing family a member 2/parotid secretory protein is an early biomarker of AKI. *J. Am. Soc. Nephrol.* 28, 3473–3478. doi: 10.1681/ASN.2016121265

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, J., Bauer, M., Moe, B., Leslie, E. M., and Li, X.-F. (2017). Multidrug resistance protein 4 (MRP4/ABCC4) protects cells from the toxic effects of halobenzoquinones. *Chem. Res. Toxicol.* 30, 1815–1822. doi: 10.1021/acs.chemrestox.7b00156

Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., et al. (2014). Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* 31, 3324–3343. doi: 10.1093/molbev/msu264

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Miao, Q., Huang, S.-Y., Qin, S.-Y., Yu, X., Yang, Y., Yang, J.-F., et al. (2015). Genetic characterization of *Toxoplasma gondii* in Yunnan black goats (Capra hircus) in southwest China by PCR-RFLP. *Parasites Vect.* 8:57. doi: 10.1186/s13071-015-0673-0

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Perumbakkam, S., Muir, W. M., Black-Pyrkosz, A., Okimoto, R., and Cheng, H. H. (2013). Comparison and contrast of genes and biological pathways responding to Marek's disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. *BMC Genomics* 14:64.

Rahmatalla, S. A., Arends, D., Reissmann, M., Said Ahmed, A., Wimmers, K., Reyer, H., et al. (2017). Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits. *BMC Genet.* 18:92. doi: 10.1186/s12863-017-0553-z

Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi: 10.1073/pnas.1217149109

Spínola, H., Lemos, A., Couto, A. R., Parreira, B., Soares, M., Dutra, I., et al. (2016). HLA-DQA1 and HLA-DQB1 allele diversity and its extended haplotypes in Madeira Island (Portugal). *Int. J. Immunogenet.* 44, 27–31. doi: 10.1111/iji.12300

Springer, N. M., and Stupar, R. M. (2007). Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* 19, 2391–2402. doi: 10.1105/tpc.107.052258

Tian, Z., Liu, G., Yin, H., Luo, J., Guan, G., Luo, J., et al. (2013). Discrimination between ovine *Babesia* and *Theileria* species in China based on the ribosomal protein S8 (RPS8) gene. *Vet. Parasitol.* 197, 354–359. doi: 10.1016/j.vetpar.2013.04.033

Valenzuela-Ponce, H., Alva-Hernández, S., Garrido-Rodríguez, D., Soto-Nava, M., García-Téllez, T., Escamilla-Gómez, T., et al. (2018). Novel HLA class I associations with HIV-1 control in a unique genetically admixed population. *Sci. Rep.* 8:6111. doi: 10.1038/s41598-018-23849-7

von Heckel, K., Stephan, W., and Hutter, S. (2016). Canalization of gene expression is a major signature of regulatory cold adaptation in temperate *Drosophila melanogaster*. *BMC Genomics* 17:574. doi: 10.1186/s12864-016-2866-0

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Wei, J., and Hemmings, G. P. (2003). TNXB locus may be a candidate gene predisposing to schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 125B, 43–49. doi: 10.1002/ajmg.b.20093

Yuan, C., Guo, T.-T., Liu, J.-B., Yue, Y.-J., and Yang, B.-H. (2017). "Conservation and utilization of indigenous goats and breeding of new breeds in China," in *Sustainable Goat Production in Adverse Environments: Volume I: Welfare, Health and Breeding,* eds J. Simões and C. Gutiérrez (Cham: Springer International Publishing), 457–472.

Zhao, Y., Zhang, J., Zhao, E., Zhang, X., Liu, X., and Zhang, N. (2011). Mitochondrial DNA diversity and origins of domestic goats in Southwest China (excluding Tibet). *Small Ruminant Res.* 95, 40–47. doi: 10.1016/j.smallrumres.2010.09.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Cao, Xu, Li, Gao, Chen, Luo and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 March 2019
doi: 10.3389/fgene.2019.00210

Detection of Co-expressed Pathway Modules Associated With Mineral Concentration and Meat Quality in Nelore Cattle

*Wellison J. S. Diniz[1,2,3], Gianluca Mazzoni[4], Luiz L. Coutinho[5], Priyanka Banerjee[2], Ludwig Geistlinger[3,6], Aline S. M. Cesar[5], Francesca Bertolini[7], Juliana Afonso[1], Priscila S. N. de Oliveira[3], Polyana C. Tizioto[5], Haja N. Kadarmideen[2] and Luciana C. A. Regitano[3]**

[1] Department of Genetics and Evolution, Federal University of São Carlos, São Carlos, Brazil, [2] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark, [3] Embrapa Pecuária Sudeste, Empresa Brasileira de Pesquisa Agropecuária, São Paulo, Brazil, [4] Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark, [5] Department of Animal Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil, [6] Graduate School of Public Health and Health Policy, The City University of New York, New York, NY, United States, [7] Department of Aquaculture, Technical University of Denmark, Kongens Lyngby, Denmark

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:
Martin Johnsson,
Swedish University of Agricultural Sciences, Sweden
Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil

***Correspondence:**
Luciana C. A. Regitano
luciana.regitano@embrapa.br

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 03 December 2018
Accepted: 26 February 2019
Published: 13 March 2019

Citation:
Diniz WJS, Mazzoni G, Coutinho LL, Banerjee P, Geistlinger L, Cesar ASM, Bertolini F, Afonso J, de Oliveira PSN, Tizioto PC, Kadarmideen HN and Regitano LCA (2019) Detection of Co-expressed Pathway Modules Associated With Mineral Concentration and Meat Quality in Nelore Cattle. Front. Genet. 10:210. doi: 10.3389/fgene.2019.00210

Meat quality is a complex trait that is influenced by genetic and environmental factors, which includes mineral concentration. However, the association between mineral concentration and meat quality, and the specific molecular pathways underlying this association, are not well explored. We therefore analyzed gene expression as measured with RNA-seq in *Longissimus thoracis* muscle of 194 Nelore steers for association with three meat quality traits (intramuscular fat, meat pH, and tenderness) and the concentration of 13 minerals (Ca, Cr, Co, Cu, Fe, K, Mg, Mn, Na, P, S, Se, and Zn). We identified seven sets of co-expressed genes (modules) associated with at least two traits, which indicates that common pathways influence these traits. From pathway analysis of module hub genes, we further found an over-representation for energy and protein metabolism (AMPK and mTOR signaling pathways) in addition to muscle growth, and protein turnover pathways. Among the identified hub genes *FASN, ELOV5*, and *PDE3B* are involved with lipid metabolism and were affected by previously identified eQTLs associated to fat deposition. The reported hub genes and over-represented pathways provide evidence of interplay among gene expression, mineral concentration, and meat quality traits. Future studies investigating the effect of different levels of mineral supplementation in the gene expression and meat quality traits could help us to elucidate the regulatory mechanism by which the genes/pathways are affected.

Keywords: AMPK pathway, co-expression analysis, intramuscular fat, RNA sequencing, tenderness

Abbreviations: AMPK, AMP-activated protein kinase; CPM, Counts *per* million; ECM, Extra Cellular Matrix; IMF, Intramuscular Fat Content; ME, Module eigengene; MM, Module Membership; QC, Quality Control; WBSF7, Warner-Bratzler Shear Force after 7 days of meat aging; WGCNA, Weighted Gene Co-expression Network Analysis.

INTRODUCTION

Meat is an important source of nutrients in the human diet. Meat quality traits such as intramuscular fat content (IMF), mineral concentration, and fatty acid profile influence consumer purchase decision (Ahlberg et al., 2014; Mateescu, 2014) and human health (Pighin et al., 2016). Mineral deficiency, mainly iron and zinc (Ritchie and Roser, 2018), and protein deficiency (Clugston and Smith, 2002), have been reported as worldwide health hazards. In addition, IMF, meat pH, and muscle mineral concentration also affect meat tenderness, flavor, and juiciness, which are major sensory traits related to eat satisfaction (Engle et al., 2000; Ahlberg et al., 2014; Pannier et al., 2014).

Brazil is one of the largest exporters of meat and meat products, and the Brazilian cattle herd is mainly composed of Nelore and its crosses (ABIEC, 2018). Despite being well adapted to tropical climate, Nelore cattle has typically less tender and marble meat when compared with European breeds due to several genetic and environmental factors (Cesar et al., 2015; Tizioto et al., 2015). Genome-wide association (GWAS) of SNPs (Tizioto et al., 2013, 2015; Cesar et al., 2014) and copy-number variations (CNVs) (Silva et al., 2016) in conjunction with transcriptomic studies (Diniz et al., 2016; Silva-Vignato et al., 2017; Geistlinger et al., 2018; Gonçalves et al., 2018), have illustrated the genetic factors affecting complex traits in Nelore. However, growing evidence suggested interplay among gene expression, mineral concentration, and meat quality traits, which are still unclear.

Multi-omic data integration has been useful to reveal potential causal and regulatory mechanisms underlying complex animal production, reproduction and welfare traits [reviewed in Suravajhala et al. (2016)]. Integrating genomic, transcriptomic, and phenotype data has contributed to an improved understanding of complex traits by identifying regulatory candidate genes and biological functions (Ponsuksili et al., 2013; Cesar et al., 2018; Geistlinger et al., 2018; Gonçalves et al., 2018). Based on that, Mateescu et al. (2017) carried out a GWAS combined with gene network analysis for association with the carcass, meat quality traits and mineral concentration. Among the identified pathways, the authors pointed out calcium-related processes, apoptosis, and TGF-beta signaling involved with these traits.

Genome-wide association and differential gene expression analyses have been fruitful in investigating the role of genes in complex phenotypes. However, biological systems are a result of complex interactions among genes and multiple regulatory mechanisms, which are not explored in the above-mentioned studies. To address the relationship between transcriptome and traits, co-expression networks have been successfully employed. This approach allows to identify and cluster highly connected genes and associate them to the phenotypes, shedding light on the common pathways underlying these traits as well as the main regulators (Langfelder and Horvath, 2008). To date, there is no information about this approach integrating meat quality traits and mineral concentration in beef cattle. In addition, we still have a lack of knowledge about the interplay among gene expression, mineral concentration, and meat quality traits. Thus, to explore regulatory pathways, putative gene regulators, and to study their relationship with muscle and mineral metabolism in Nelore skeletal muscle, we integrated gene expression, eQTL variation, mineral concentration (macro and micro minerals), and meat quality traits (intramuscular fat, shear force, and meat pH) based on a network approach.

MATERIALS AND METHODS

Ethics Statement

The Institutional Animal Care and Use Committee (IACUC) from the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA – Pecuária Sudeste) approved all experimental procedures involving the animals used in this study.

Animals and Phenotyping

A total of two hundred Nelore steers (produced at Embrapa Pecuária Sudeste, São Carlos – Brazil) were used in this study. The experimental design, production system, and animal management were previously described (Tizioto et al., 2015; Diniz et al., 2016). Briefly, animals were raised in the grazing system until 21 months of age when they were taken to three feedlots under similar nutritional and sanitary management. The Nelore steers with an average age of 25 months were harvested at commercial facilities after about 90 days of feeding and the *Longissimus thoracis* (LT) muscle samples were collected.

The steaks (2.5 cm) harvested as a cross-section of the LT muscle (11th and 13th ribs) collected at slaughter were used to measure the beef quality traits as described (Tizioto et al., 2013; Cesar et al., 2014). The traits evaluated were tenderness (Warner-Bratzler shear force – WBSF7, kg) measured 7 days after slaughter, meat pH measured 24 h after slaughter along with intramuscular fat (IMF%) (Tizioto et al., 2013).

Tissue samples were used for total RNA extraction (Diniz et al., 2016) and mineral measurement (Tizioto et al., 2014). The concentration of macro minerals [calcium (Ca), magnesium (Mg), phosphorus (P), potassium (K), sodium (Na), sulfur (S)] and micro minerals [chromium (Cr), cobalt (Co), copper (Cu), manganese (Mn), selenium (Se), iron (Fe), and zinc (Zn)] were measured using inductively coupled plasma-optical emission spectrometry (ICP OES; Vista Pro-CCD ICP OES1, radial view, Varian, Mulgrave, Australia) as described by Tizioto et al. (2014).

Genome Expression Profile, Sequencing, and Data Processing

The LT muscle samples were collected immediately after slaughter, snap frozen in liquid nitrogen and kept at −80°C until RNA extraction. To extract RNA, approximately 100 mg of frozen tissue was used, and total RNA was purified using Trizol® standard protocol (Life Technologies, Carlsbad, CA, United States). The mRNA concentration and quality were evaluated in the Bioanalyzer 2100® (Agilent, Santa Clara, CA, United States).

The Illumina TruSeq® RNA Sample Preparation Kit v2 Guide (San Diego, CA, United States) protocol was used to generate

cDNA libraries for each sample using 2 μg of total RNA as input. Library preparation and sequencing were conducted by ESALQ Genomics Center (Piracicaba, São Paulo, Brazil). cDNA libraries were purified and validated using Agilent 2100 Bioanalyzer (Santa Clara, CA, United States). Paired-end (PE) sequencing was performed on Illumina Hiseq 2500® (San Diego, CA, United States) platform following the standard protocols. The samples were multiplexed and run on multiple lanes to obtain 2 × 100 bp reads.

The PE reads were filtered using the Seqyclean package version 1.4.13 ([1]Zhbannikov et al., 2017), which removed all reads with a mean quality under 24, length under 65 bp, as well as the adapter sequences. Quality control (QC) of raw RNA-Seq reads was carried out with FastQC version 0.11.2 ([2]Andrews, 2010) and MultiQC version1.4 ([3]Ewels et al., 2016).

Read mapping and gene counting were carried out by STAR aligner version 2.5.4b (Dobin et al., 2013) using a reference genome (*Bos taurus,* ARS-UCD1.2) and gene annotation file (release 106) obtained from NCBI (NCBI, 2018). One sample with mapping rate lower than 70% was removed out for further analyses.

The data editing was done using the Bioconductor package edgeR version 3.20.9 (Robinson et al., 2010). Taking into account that low expressed genes are less reliable and indistinguishable from sampling noise (Tarazona et al., 2015), the read counts per gene were normalized to counts per million (*cpm* function). The genes with less than one *cpm* in more than 90% of the samples were filtered out. Gene counts were normalized applying the variance stabilizing transformation (*VST*) from DESeq2 version 1.18.1 (Anders and Huber, 2010).

Potential biases due to technical variation in gene expression among samples were evaluated by applying a Principal Component Analysis (PCA) and hierarchical clustering on normalized data using NOISeq version 2.22.1 (Tarazona et al., 2015). A linear model was fitted in order to adjust the gene expression matrix for batch effect (flow cell). To this end, the *removeBatchEffect* function from Limma (version 3.34.9) R package (Ritchie et al., 2015) was used. Three samples were identified as outliers. Thus, 12 known housekeeping genes were selected based on the literature (*ACTB, API5, EIF2B2, GAPDH, GUSB, HMBS, PGK1, PPIA, RPL13A, VAPB, YWHAZ*) to evaluate their variability on the samples. The housekeeping genes expression confirmed these samples as outliers, and therefore, they were filtered out.

Network Gene Co-expression Analysis

A co-expression approach was applied using the WGCNA R package version 1.63 (Langfelder and Horvath, 2008). The method adopted for constructing the networks included two steps: First, a similarity co-expression network was calculated with Pearson's correlation for all genes, followed by transformation to a signed adjacency matrix (AM) by using the soft thresholding power β, to which co-expression similarity is raised. Based on the criteria of approximating scale-free topology, we chose the power of β = 12 such that the resulting network satisfies the scale-free topology (linear regression model fitting index $R^2 = 0.80$).

Outlier animals ($n = 2$) were identified based on hierarchical clustering and filtered out (as they had a lower number of counts compared to other samples) after WGCNA quality control, as suggested by the WGCNA authors. Accordingly, 194 animals and 11,996 genes were used to construct an undirected, signed network. Topological overlap measure (TOM) was computed from AM where TOM was converted to dissimilarity TOM. Based on TOM dissimilarity, we used the dynamic tree cut v.1.63.1 (Langfelder et al., 2008) to identify the modules as the branches of the resulting dendrogram. As parameters, the minimum size per module was set to 50 genes with a high sensitivity to split the clusters (deepSplit = 4). Genes with a similar expression pattern across samples were grouped into the same module and arbitrarily labeled by number.

Weighted Gene Co-expression Network Analysis was used for summarizing the obtained modules by a concept of eigengene. Eigengenes are the first principal component of the expression matrix for each module and represent the weighted average of expression profile for each module. Modules highly correlated were merged based on the ME dissimilarity threshold of 0.2 leading to the final set of modules for constructing the network.

Trait Association Analysis and Module Selection

After the phenotypic data were mean-centered and scaled, a linear model was fitted to analyze the association between the expression profiles of the MEs and the phenotypes (Li et al., 2018). The model included the place of birth, the season of production, and animal's age, according to the equation:

$$y_{ijkl} = \mu + C_i + G_j + A_k + T_l + \varepsilon_{ijkl}$$

Where:

y_{ijkl}: is the expression level of the eigengene in each module ($n = 23$);

μ : is the intercept of ME;

C_i : is the fixed effect for the place of birth (three levels = CPPSE, IMA, NOHO);

G_j: is the fixed effect for the season of production (three levels = 2009, 2010, 2011);

A_k : is the covariate for the animal's age, in days;

T_l : is the trait observation for each animal;

ε_{ijkl}: is the random residual effect associated with each observation.

Modules associated with at least two beef quality or mineral traits ($p \leq 0.05$) were selected for further analyses.

Pathway Over-Representation Analysis

Pathway analysis was performed using ClueGO version 2.5.1 to identify gene KEGG pathways over-represented in the selected modules (Bindea et al., 2009). Redundant terms were grouped based on the kappa score = 0.4 (Bindea et al., 2009). The *p*-value was calculated and corrected with a Bonferroni step down. Only

[1]https://github.com/ibest/seqyclean

[2]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[3]http://multiqc.info/

pathways with a *p*-value (pV) $p \leq 0.05$ were selected. These analyses were carried out based on the *B. taurus* annotation, and the network visualization was performed on Cytoscape version 3.6.1 (Shannon et al., 2003).

Hub Gene Selection

Highly connected genes (hub genes) are supposed to be the main regulators in the network and have a pivotal biological role concerning the associated trait (Langfelder and Horvath, 2007, 2008). Hub genes in the associated modules were selected based on the module membership ≥ 0.8 (Langfelder and Horvath, 2008). Among them, hub genes partaking in over-represented biological pathways previously identified were retained. Moreover, over-representation pathway analysis including all hub genes was applied following the approach previously described.

Integration of eQTL and Co-expression Modules

A list of eQTLs from the same population and dataset (Cesar et al., 2018) evaluated in this work was provided. The dataset included 1,268 *cis*- and 10,334 *trans*-eQTLs based on the association between 461,466 SNPs and the expression level of 11,808 genes from 192 animals. Since the eQTLs have a known effect on gene expression, the eQTLs that target the hub genes (MM ≥ 0.8) in the selected modules were evaluated. A Fisher's exact test was applied to assess the module under/over-representation (FDR = 0.05).

RESULTS

We applied a network-based approach to identify relevant genes and pathways associated with meat quality and mineral concentration in Nelore cattle (**Figure 1**). Based on the transcriptomic profiles of skeletal muscle samples of 194 steers, we constructed a signed weighted gene co-expression network with WGCNA (Langfelder and Horvath, 2008). From co-expressed modules and pathway analysis, we thereby identified several hub genes significantly associated with meat quality traits and mineral concentration.

Descriptive Statistics and Correlation Estimates

We analyzed gene expression levels as measured with RNA-seq for association with three meat quality traits (intramuscular fat, meat pH, and tenderness) and the concentration of 13 minerals (Ca, Cr, Co, Cu, Fe, K, Mg, Mn, Na, P, S, Se, and Zn) available for a varying number of samples (ranging from 57 to 194, **Supplementary Table S1**). The genetic variance and heritability for the traits evaluated here, obtained from this population, ranged from low to moderate as previously published (Tizioto et al., 2013, 2015). A summary of descriptive statistics for each trait is in **Supplementary Table S1** and **Figure 2**.

We performed clustering analysis to identify similarities between traits (**Figure 3** – top). We identified four clusters as follows: cluster 1 (WBSF7 and Cr), cluster 2 (Co, Cu, Mn, and IMF), cluster 3 (Fe, Ca, S, Zn, Na, P, Mg, and K), and cluster 4 (pH and Se). The pair-wise correlation within all traits is provided in **Supplementary Figure S1**. Significant and strong correlation ranged from 0.45 to 0.99 among minerals in the cluster 3 ($p \leq 0.05$). We identified positive correlation among IMF with some minerals (Ca = 0.25, Cu = 0.23, Mn = 0.24, K = 0.17, Na = 0.3, S = 0.18, and Zn = 0.23) ($p \leq 0.05$). Meat pH was positively correlated with Se ($r = 0.29$), whereas negatively associated with Fe (−0.17), Mg (−0.22), P (−0.25), K (−0.21), Na (−0.26), S (−0.17), and Zn (−0.22) (**Supplementary Figure S1**). No significant correlation was observed between tenderness (WBSF7), IMF, and meat pH.

Data Processing and Co-expression Network Construction

On average, a total of 13 million of 100 bp paired-end reads per sample were generated. Around 96.71% of unique reads were mapped to the reference *B. taurus* genome (ARS-UCD1.2). Taking into account that low expressed genes are less reliable and indistinguishable from sampling noise (Tarazona et al., 2015), we filtered out the genes with less than one *cpm* in more than 90% of the samples. In addition, four samples were removed because they had a mapping rate lower than 70% or showed high variability on the housekeeping genes expression (see methods). Thus, we used 11,996 genes and 194 samples for the co-expression analysis.

Considering the WGCNA assumptions, the weighted network starts from the level of thousands of genes, identifies modules of co-expressed genes, summarizes the module expression profile as the first principal component (ME), and relates the MEs with the trait of interest (Langfelder and Horvath, 2008). The MM value quantifies the degree of co-expression of a gene with other genes within a module, thereby enabling the identification of intramodular hub genes.

From clustering 11,996 genes with WGCNA, we obtained 23 modules labeled by number (**Figure 3**). The module size ranged from 69 genes (M9) to 2,008 genes (M14) (**Figure 3** – bottom). The proportion of variance explained by the eigengenes ranged from 0.18 (M20) to 0.53 (M5) (**Supplementary Table S2**).

Trait Association and Pathway Enrichment Analysis

We performed an association analysis to identify the relationship between network and traits. This analysis measures the strength of the effect and the direction of the association between the module (eigengenes) and the trait. Thus, if the association is positive, it means the trait increases with increasing "eigengene expression" or vice-versa. We selected seven modules (M1, M5, M6, M7, M8, M9, and M17), associated with at least two traits ($p \leq 0.05$) (**Figure 3** – bottom) once we also want to point out shared pathways among traits. We found the highest number of significant associated modules between M5 (ten associations; negative with IMF, and the concentration of Mn, Fe, Ca, S, Zn, Na, P, Mg, and K), followed by M8

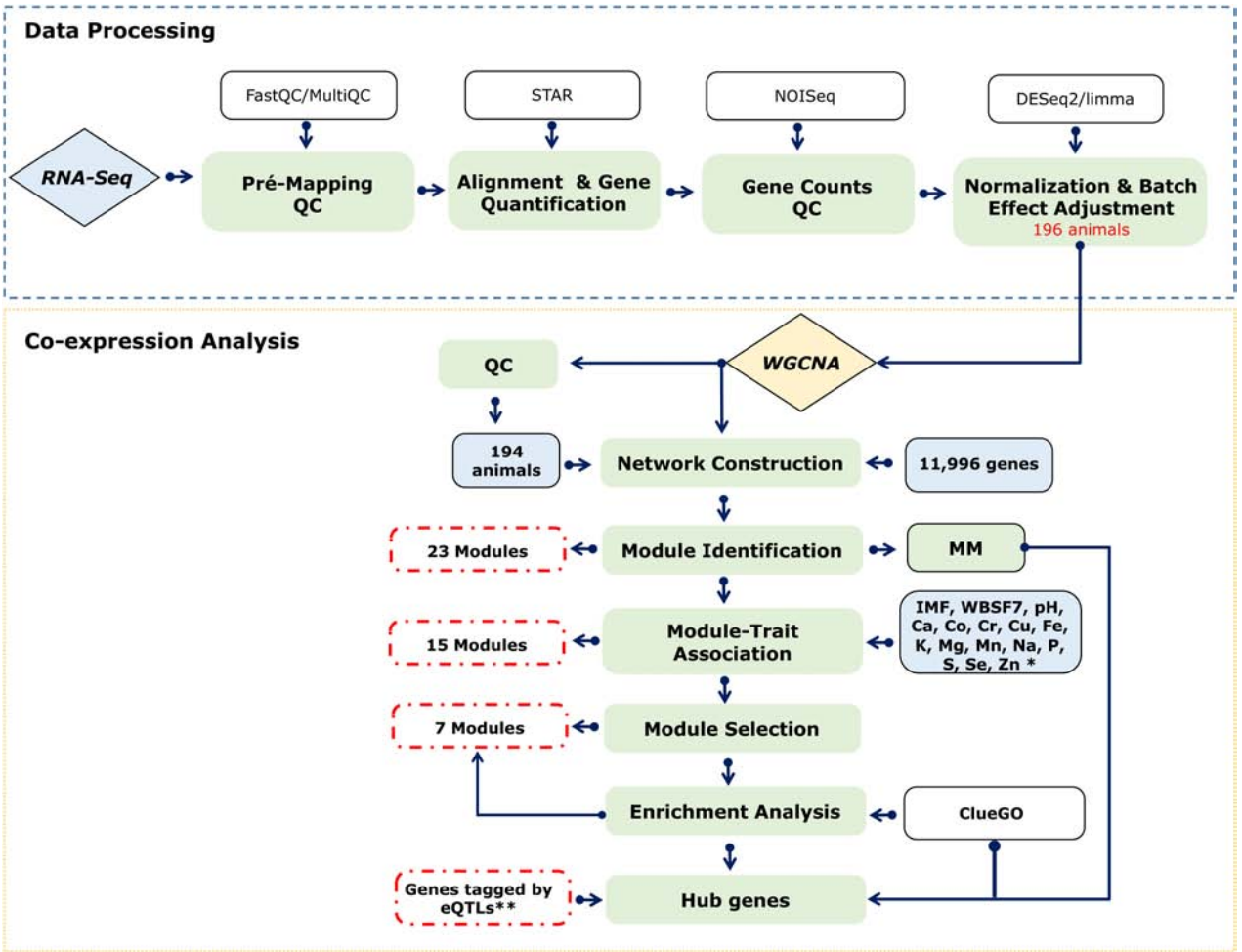


FIGURE 1 | Workflow. Green boxes represent the main analysis steps that were involved in data processing and co-expression analysis. Tools applied in each step are shown in white boxes. Inputs and outputs are shown in blue and dashed red boxes, respectively. *A varying number of samples was analyzed for each trait (**Supplementary Table S1**). **Data from an eQTL analysis carried out for the same population used in this study (Cesar et al., 2018).

(nine associations; positive with Cr, negative with IMF, and the minerals of cluster 3, except Zn). The average expression profile of M17 module showed association with three traits (positive with WBSF7, Co, and Mn) along with M7 (negative with Na and IMF, and positive with Cr concentration). For the modules M1, M6, and M9, we found an association with two traits. We identified a positive association among M6 and M9 with Cr concentration while a negative association was observed between M9 with IMF, and M6 with Fe concentration. M1 was positively associated with the concentration of Cr and Co. The modules with none or only one trait association were not included for further analysis.

The module membership values for all the genes for selected modules are given in **Supplementary Table S3**. We carried out a pathway over-representation analysis on ClueGo version 2.5.1 for the seven selected modules (**Table 1** and **Supplementary Table S4**) to identify meaningful metabolic pathways involved with meat quality traits and mineral concentration. We detected several pathways (p-Value $\leq$ 0.05, group p-value corrected with Bonferroni step down) mainly related to energy and protein metabolism, such as AMPK and mTOR signaling pathways.

Hub Gene Selection, Pathway Analysis, and Integration With eQTLs

Highly connected genes are likely to play an important role both in the network's topology and biological pathways. In this way, we combined a pathway-based gene analysis for each selected module (**Supplementary Table S4**) and gene connectivity measure (MM $\geq$ 0.8) (**Supplementary Table S3**) selecting 82 hub genes (**Table 1**, see methods). Further, taking advantage of an eQTL study carried out in the same population (Cesar et al., 2018), we screened whether the genes in the modules were underlying eQTLs, and applied a Fisher's exact test to assess the module over-representation. We identified 323 genes targeted by 760 unique eQTLs (**Table 1** and **Supplementary Table S5**) into the seven modules. In addition, we identified 24 out of 323 genes with a

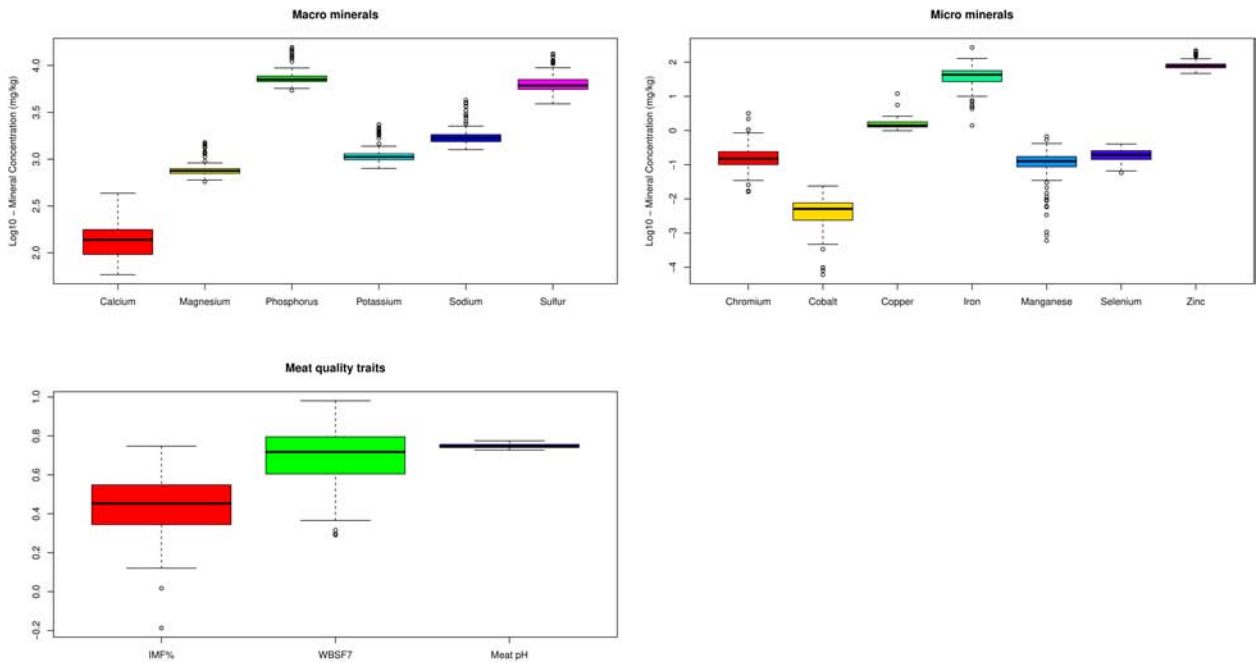


FIGURE 2 | Box plot of mineral concentration (macro and micro) and meat quality traits. IMF% – Percentage of intramuscular fat content; WBSF7 – Warner-Bratzler Shear Force after 7 days of meat aging; The data are in *log10* scale.

MM $\geq$ 0.8, and six of them are part of the hub gene list (**Table 1**). However, based on the Fisher's exact test (FDR = 0.05) no significant over-/under-representation was detectable in these modules.

To gain further insights into their functions as well as to integrate the pathways among the modules, we carried out a KEGG pathway analysis. Considering a kappa score = 0.4 and *p*-Value $\leq$ 0.05 (**Figure 4** and **Supplementary Table S6**), we clustered the identified pathways into eight groups. The pathways related to energy metabolism were clustered together and included AMPK, peroxisome proliferator-activated receptors (PPAR), insulin, glucagon, and adipocytokine signaling pathways. We also identified ubiquitin-mediated proteolysis and biosynthesis of fatty acids pathways over-represented in this network.

DISCUSSION

In this study, we analyzed genome-wide co-expression in skeletal muscle for association with mineral concentration and meat quality traits. Skeletal muscle metabolism is an integrated system dependent on the efficient coordination of gene expression, which are tightly regulated (Smith et al., 2013). We found several co-expression modules associated with two or more minerals, meat tenderness, and IMF, which indicates that common pathways influence these traits. From pathway analysis of module hub genes, we further found an over-representation for energy and protein metabolism (AMPK and mTOR). These pathways have been reported as the main drivers regulating energy balance in muscle (Smith et al., 2013). AMPK and mTOR are metabolically linked, nutritional and hormonal responsive, with an intricated relationship with insulin, thyroid hormone (TH), and TGF-beta signaling pathways (Xu et al., 2012), which were reported here as well. In addition, these pathways have been associated with muscle development, fat deposition, and beef quality traits (Du et al., 2009). Pathways related with muscle structure such as extracellular matrix, and focal adhesion, identified here, have also been identified in cattle co-expression networks (Reverter et al., 2006). The above-mentioned pathways are not the only ones acting on muscle metabolism. However, they showed an interaction with mineral concentration and meat quality in our study.

Phenotype Correlation and Co-expression Network Analysis

In agreement with previous reports, we found several minerals positively correlated with IMF, but negatively correlated with meat pH. For instance, Cu-supplemented Angus were found with reduced back fat and reduced serum cholesterol level (Engle et al., 2000). Pigs supplemented with Mn showed an increased marbling and decreased pH consistent with the correlation identified here (Constantino et al., 2014). Furthermore, Se supplementation improved pork meat quality traits by increasing muscle pH (Calvo et al., 2017). In addition, these studies reported a positive effect against lipid oxidation. On the other hand, reduced levels of IMF were associated with low Zn concentration in lambs (Pannier et al., 2014).

Co-expression analysis resulted in 23 modules from which we considered seven modules for further analysis based on their

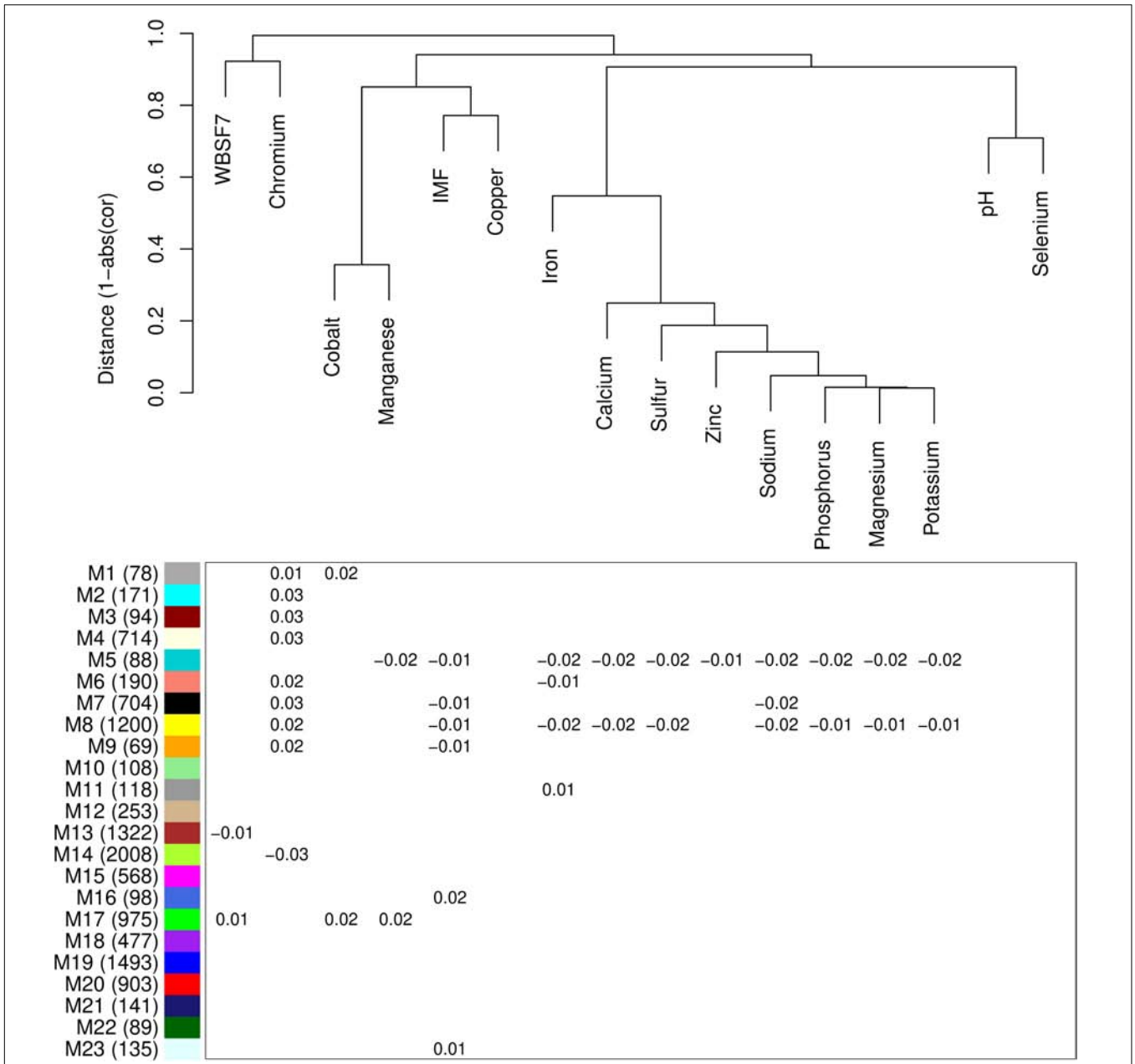


FIGURE 3 | Hierarchical clustering of phenotypic correlation between traits (top) and module-trait association analysis (bottom). Modules are labeled by number on the *y*-axis with the number of contained genes in parenthesis. Each column represents a trait as indicated on the corresponding dendrogram branch. For significantly associated modules ($p \leq 0.05$), the coefficient from the linear model is given within the cell.

association with at least two traits. The genes in modules like M5, M7, M8, and M9 were associated with IMF and several minerals suggesting a certain extent of co-regulation. It's well known that minerals are essential in a wide range of biological processes. Here, we provide evidences that mineral content and meat quality traits are interrelated, as well as interplay with specific genes and pathways (as discussed below).

Variation in eQTL loci can explain a substantial fraction of variation observed on the gene expression level (Wang and Michoel, 2016). It has been observed that variation in eQTL loci is associated with concerted expression changes of many genes in co-expression clusters, thereby also impact the phenotype. Screening the detected co-expression modules, we found 323 genes affected by at least one eQTL. Despite 132 eQTLs targeting more than one module, most of the eQTLs were module-specific. However, no significant over-/under-representation (Fisher's exact test) was detectable in these modules, suggesting that other regulatory mechanisms are involved. Despite that, the expression level of six hub genes was found affected by *trans*-eQTLs. These genes are involved

TABLE 1 | Module characterization.

Module[a]	eQTLs[b]	TGE[c]	Hub genes[d]	Enriched pathways[e]
M1 (78)	13	8 (4)	(10) *RSAD2, LOC100139670, MX1, IFIH1, OAS1Y, DDX58, EIF2AK2, IRF7, LOC512486, IFI16*	• NOD-like receptor signaling pathway
M5 (88)	45	15 (3)	(9) *PTPRC, CTSS,* ***DOCK2***, ***IL10RA***, *FERMT3, ITGB2,* ***LCP2***, *CORO1A, PTPRJ*	• Phagosome • Cell adhesion molecules (CAMs)
M6 (190)	89	20 (6)	(20) *GNAI1, PLIN1, FABP4, PPP1R1B, PCK2, ADIPOQ,* ***PDE3B***, *TKT, MGST1, LIPE, ELOVL6, MRAS, ACACA, GNG2, HACD2,* ***FASN***, *FBP1,* ***ELOVL5***, *PCK1, G6PD*	• Adipocytokine signaling pathway • AMPK signaling pathway
M7 (704)	276	76 (2)	(11) *CAV1, COL4A2, GNAI2, SEPT9, TNFRSF1A, YWHAB, COL4A1, TMSB4X, SPTAN1, MSN, PARVA*	• Ras signaling pathway • Focal adhesion
M8 (1,200)	414	129 (5)	(10) *COL1A2, DCN, FN1, COL1A1, FGFR1, COL3A1, CD44, ITGB5, DSE, CTSK*	• Glycosaminoglycan degradation Steroid biosynthesis
M9 (69)	16	9 (2)	(1) *JUND*	• TGF-beta signaling pathway • Osteoclast differentiation
M17 (975)	126	66 (2)	(21) *ASH1L, BIRC6, MED13, HERC1, MED13L, KMT2A, HUWE1, KMT2C, ATM, HERC2, KMT2D, MED1, GNAQ, SMC1A, TPR, NSD1, UBR5, ARHGAP5, KAT6A, PRDM2, ITCH*	• Apoptosis • mTOR signaling pathway
Total – 3304	**979[f]**	**323 (24)**	**82**	

The table shows hub genes and eQTL information for each module found to be significantly associated with two or more traits in ***Figure 3****.* [a]*Selected modules with the number of contained genes in parenthesis.* [b]*eQTLs – Number of eQTLs associated with genes in a module (Based on Cesar et al., 2018);* [c]*TGE – Number of module genes associated with eQTLs. In the parenthesis are the number of genes with a MM ≥ 0.8;* [d]*Selected hub genes based on pathway analysis and MM; Hub genes associated with eQTLs are in bold;* [e]*Pathways from module over-representation analysis taking all genes into the module (****Supplementary Table S4****) Group p-Value ≤ 0.05;* [f]*760 unique eQTLs identified.*

with lipid metabolism [fatty acid synthase (*FASN*), and ELOVL fatty acid elongase 5 (*ELOVL5*), phosphodiesterase 3B (*PDE3B*)], immune system [lymphocyte cytosolic protein 2 (*LCP2*), and interleukin 10 receptor subunit alpha (*IL10RA*)], and actin remodeling (dedicator of cytokinesis 2 – *DOCK2*).

Pathway Analysis

Over-representation pathway analysis in the selected modules (**Table 1** and **Supplementary Table S4**) yielded glycosaminoglycan biosynthesis and degradation, lysosome, and steroid biosynthesis in the M8 module. Phagosome, cell adhesion molecular pathways, and NOD-like receptor signaling pathway were found enriched in M1 and M5. For the M17, enriched pathways included protein synthesis pathways such as mTOR, PI3K-Akt, TH, and AMPK signaling. We also found protein degradation pathways enriched in the M17 module such as ubiquitin-mediated proteolysis. TGF-beta signaling and osteoclast differentiation were enriched in the M9 module. Energy metabolism pathways were found enriched in M6, including glycolysis, fatty acid biosynthesis, AMPK, and insulin signaling. Ras, PI3K-Akt signaling pathways, and protein processing were found enriched in M7.

We also carried out cross-module enrichment analysis considering all hub genes, which indicated that the AMPK signaling pathway plays an important role for muscle mineral metabolism and meat quality traits. The genes of the AMPK pathway were also associated with IMF, Cr, and Fe. Furthermore, the AMPK pathway was also found enriched in genes of M17 (associated with WBSF7, Co, and Mg) and M6 (associated with Cr and Fe concentration).

Energy and Lipid Metabolism

AMP-activated protein kinase signaling is a major regulator of the cellular energy status, protein metabolism, and muscle metabolism (Je et al., 2006; Du et al., 2009; Mihaylova and Shaw, 2011). We found carbohydrate and fatty acid metabolism connected by the AMPK pathway (**Figure 4**). Hub gene *ACACA* was thereby involved in pyruvate metabolism, glucagon and insulin signaling pathways. Co-expressed in the M6 module, *ACACA* and *FASN* encode rate-limiting enzymes for long-chain fatty acid synthesis (Mihaylova and Shaw, 2011; Ropka-molik et al., 2017). *ACACA* catalyzes malonyl-CoA from acetyl-CoA, which is a substrate for the *FASN* enzyme in *de novo* fatty acid synthesis (Menendez and Lupu, 2007; Du et al., 2009). These genes, as well as fatty acid binding protein 4, adipocyte (*FABP4*), are regulated by the thyroid hormone responsive gene (*THRSP*) (Graugnard et al., 2009; Loor, 2010; Oh et al., 2014).

The co-expression of these genes, as well as the negative association between Fe concentration and lipid metabolism, were reported in our previous RNA-Seq work where *FASN*, *THRSP,* and *FABP4* were shown to be downregulated in animals with low Fe concentration in muscle (Diniz et al., 2016). Hay et al. (2016) reported a major role of Fe for lipid oxidative metabolism based on the downregulation of peroxisome proliferator-activating receptor gamma coactivator 1α *(PPARG1A)* measured by qRT-PCR. TH is also essential for energy metabolism regulation, and Fe deficiency was found to impair TH synthesis and its regulatory function (Cunningham et al., 1998). Adipogenic genes are responsive to *PPARG* and *TH* (Graugnard et al., 2009). Thus, reduced adipogenesis has been associated with Fe deficiency (Cunningham et al., 1998; Diniz et al., 2016; Hay et al., 2016).

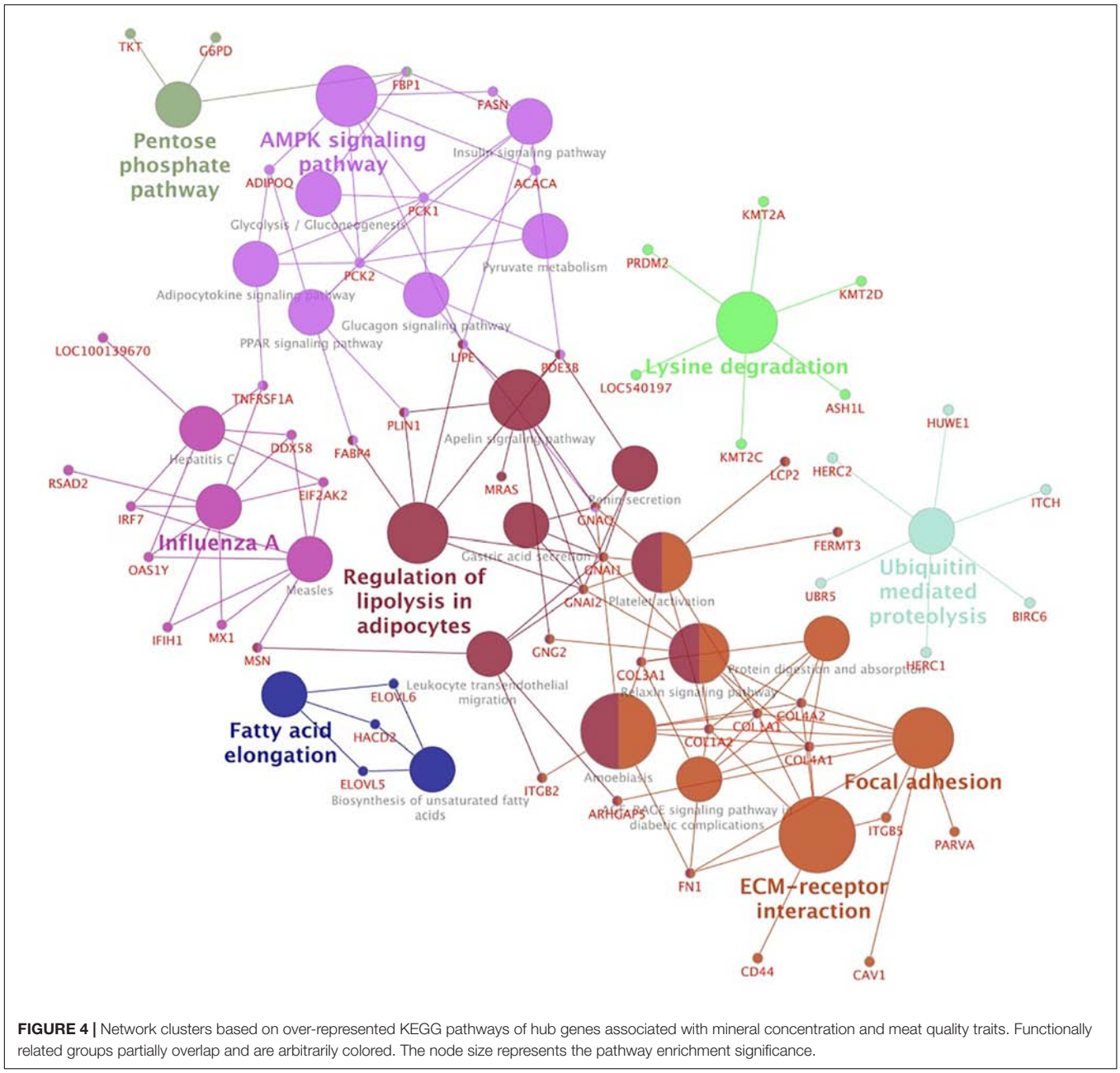


FIGURE 4 | Network clusters based on over-represented KEGG pathways of hub genes associated with mineral concentration and meat quality traits. Functionally related groups partially overlap and are arbitrarily colored. The node size represents the pathway enrichment significance.

In addition to factors that increase the intracellular cyclic AMP level (Omar et al., 2009), Cr increases AMPK activity and positively affects the insulin sensitivity in skeletal muscle cells (Hoffman et al., 2015). As part of the insulin pathway, we found phosphoenolpyruvate carboxykinase 1 and 2 (*PCK1*, *PCK2*), fructose-bisphosphatase 1 (*FBP1*), and phosphodiesterase 3B (*PDE3B*), major regulators of glycolysis and gluconeogenesis (Pilkis and Granner, 1992). The *PDE3B* enzyme is stimulated by insulin and cAMP (Degerman et al., 2011) and affects the activation of AMPK (Omar et al., 2009). AMPK activation inhibits fatty acid synthesis and gluconeogenesis via repression of *ACACA* and *PCK*, respectively (Hardie, 2011). Unlike Fe, the concentration of Cr showed a positive correlation with M6. These minerals may have an antagonistic relationship (Staniek and Wójciak, 2018). However, the correlation between Fe and Cr concentration was not significant in this study most likely due to the limited sample size for Cr concentration.

Supplementing goats with Cr decreased the expression level of *ACACA*, *FASN*, and *FABP4* (Sadeghi et al., 2015) as measured by RT-PCR. Furthermore, increased *Longissimus* muscle area and reduced fat thickness was associated with a downregulation of *ACACA* expression in Cr-supplemented goats (Najafpanah et al., 2014). It seems to follow that Cr supplementation can improve meat quality by altering the direction of energy accumulation from fat deposition toward muscle growth in goats

(Najafpanah et al., 2014; Sadeghi et al., 2015). Cr-supplemented Angus-cross steers were also found with increased *Longissimus* muscle area and decreased IMF without affecting growth performance (Kneeskern et al., 2016). Similar results were reported for Cr-supplemented pigs which showed lower backfat thickness and fat percentage (Pamei et al., 2014).

Muscle Development, Structure, and Proteolysis

As part of the TGF-beta signaling pathway, we identified the transforming growth factor beta 3 (*TGFB3*), which is involved in muscle proliferation, differentiation, and growth (Nishimura, 2015). However, muscle hypertrophy results from a balance of protein turnover in which AMPK signaling negatively affects the protein synthesis (Du et al., 2009). AMPK signaling also acts on cytoskeletal dynamics (Mihaylova and Shaw, 2011). As pointed out in **Figure 4**, common genes act on focal adhesion and ECM-receptor interaction. For these pathways, we found members of the collagen gene family (*COL1A1, COL1A2, COL3A1, COL4A1,* and *COL4A2*), glycoproteins and proteoglycans such as fibronectin 1 (*FBN1*) and decorin (*DCN*), respectively. These molecules are structural components of the ECM and are thus critical for muscle development (McCormick, 2009; Nishimura, 2015). These genes were also found associated with meat quality traits such as tenderness and IMF (Ponsuksili et al., 2013; Cesar et al., 2015; Nishimura, 2015). Except for *COL4A1* and *COL4A2*, all collagen genes reported above and which we found co-expressed in M8 were associated with the concentration of Ca, Cr, Fe, K, Mg, Na, P, S, and IMF. Tajima et al. (1981) reported that hypocalcemic fibroblast cells showed an increased synthesis of collagen. Fe concentration has also been associated with collagen metabolism due to the iron-dependent enzymes involved in collagen synthesis (Cammack et al., 1990).

We found ubiquitin-mediated proteolysis enriched across modules as well as for genes in the M17 (**Supplementary Table S3**), which was associated with WBSF7, Co, and Mn. Proteolytic enzymes are important for protein turnover and postmortem meat aging (Koohmaraie et al., 2002; Gonçalves et al., 2018). Baculoviral IAP repeat containing 6 (*BIRC6*) is a caspase inhibitor and apoptotic suppressor protein (Verhagen et al., 2001). *BIRC6* is part of the ubiquitin-mediated proteolysis pathway and was positively associated with M17. By impairing proteolysis, the up-regulation of *BIRC6* likely increases shear force (Liu et al., 2016). Genes from the E3 ubiquitin-protein ligase family (*HERC1, HERC2, HUWE1, ITCH,* and *UBR5*) were also identified in agreement with a recent report that found ubiquitination and apoptosis to be potential regulators of meat tenderness in Nelore cattle (Gonçalves et al., 2018).

CONCLUSION

We demonstrated transcriptional relationships among mineral concentration and meat quality traits in the skeletal muscle of Brazilian Nelore cattle. We identified 82 hub genes across seven co-expression modules which seem to be critical for this interplay. The AMPK and mTOR signaling pathways were hereby found to link mineral and muscle metabolism in Nelore cattle. Future studies investigating different levels of mineral supplementation, the mineral interaction, and their effect in the gene expression and meat quality traits could help us to elucidate the regulatory mechanism by which the genes/pathways are affected.

DATA AVAILABILITY

All relevant data are within the paper and its Supporting Information files. All sequencing data is available in the European Nucleotide Archive (ENA) repository (EMBL-EBI), under accession PRJEB13188, PRJEB10898, and PRJEB19421 (https://www.ebi.ac.uk/ena/submit/sra/). All additional datasets generated and analyzed during this study may be available upon request from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

WD, PT, LR, LC, and HK conceived the idea of this research. WD, GM, LG, FB, and AC, JA, PdO, PT carried out the bioinformatics and data analysis. WD, GM, PB, FB, HK, JA, LC, LR collaborated with the interpretation of results, discussion and review the manuscript. WD, PB, and GM drafted the manuscript. All authors have read and approved the final manuscript and agreed to be accountable for the content of the work.

FUNDING

This study was conducted with funding from EMBRAPA (Macroprograma 1, 01/2005), FAPESP (grant# 2012/23638-8), and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. LR and LC were granted CNPq fellowships. WD was granted by São Paulo Research Foundation (FAPESP) grant# 2015/09158-1 and grant# 2017/20761-7 scholarships. Federal University of São Carlos (PROAP/PNPD) granted funding for publishing.

ACKNOWLEDGMENTS

We are thankful to Bruno G. N. Andrade for the server management and support; Dr. Ana Rita Araújo Nogueira and Dr. Caio F. Gromboni for the mineral data; and the Technical University of Denmark (DTU Compute) for accepting the first author as a visiting scholar.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00210/full#supplementary-material

FIGURE S1 | Correlation matrix of mineral concentration and meat quality traits in Nelore cattle. Each cell displays the correlation value when significant ($p \leq 0.05$). The matrix is color-coded by correlation according to the color legend.

TABLE S1 | Summary statistics of meat quality traits and mineral concentration in Nelore cattle.

TABLE S2 | The proportion of variance explained by the module eigengene (MEs).

TABLE S3 | Gene list and module membership for each selected module. Spreadsheet tabs are divided by module.

TABLE S4 | Summary of pathway analysis from ClueGo for genes clustered into the selected modules. Spreadsheet tabs are divided by module.

TABLE S5 | Genes targeted by eQTLs for each selected module. Genes with MM $\geq$ 0.8 are highlighted in bold. Spreadsheet tabs are divided by module.

TABLE S6 | Summary of pathway analysis from ClueGo for hub genes.

REFERENCES

ABIEC (2018). *Perfil da pecuária no Brazil. São Paulo*. Available at: http://abiec.siteoficial.ws/images/upload/sumario-pt-010217.pdf

Ahlberg, C. M., Schiermiester, L. N., Howard, T. J., Calkins, C. R., and Spangler, M. L. (2014). Genome wide association study of cholesterol and poly- and monounsaturated fatty acids, protein, and mineral content of beef from crossbred cattle. *Meat Sci.* 98, 804–814. doi: 10.1016/j.meatsci.2014.07.030

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Andrews, S. (2010). *FastQC: a Quality Control Tool for High Throughput Sequence Data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi: 10.1093/bioinformatics/btp101

Calvo, L., Toldrá, F., Rodríguez, A. I., López-Bote, C., and Rey, A. I. (2017). Effect of dietary selenium source (organic vs. mineral) and muscle pH on meat quality characteristics of pigs. *Food Sci. Nutr.* 5, 94–102. doi: 10.1002/fsn3.368

Cammack, R., Wrigglesworth, J. M., and Baum, H. (1990). "Iron-dependent enzymes in mammalian systems," in *Iron:Transport and Storage*, eds P. Ponka, H. M. Schulman, R. C. Woodworth, and G. W. Richter (Boca Raton, FL: CRC Press), 17–39.

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., et al. (2015). Putative regulatory factors associated with intramuscular fat content. *PLoS One* 10:e0128350. doi: 10.1371/journal.pone.0128350

Cesar, A. S. M., Regitano, L. C. A., Mourão, G. B., Tullio, R. R., Lanna, D. P. D., Nassu, R. T., et al. (2014). Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. *BMC Genet.* 15:39. doi: 10.1186/1471-2156-15-39

Cesar, A. S. M., Regitano, L. C. A., Reecy, J. M., Poleti, M. D., Oliveira, P. S. N., De Oliveira, G. B., et al. (2018). Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. *BMC Genomics* 19:499. doi: 10.1186/s12864-018-4871-y

Clugston, G. A., and Smith, T. E. (2002). Global nutrition problems and novel foods Global nutrition problems and novel foods S101. *Asia Pac. J. Clin. Nutr.* 11, 100–111. doi: 10.1046/j.1440-6047.11.s6.1.x

Constantino, C., de Azambuja Ribeiro, E. L., Bridi, A. M., Tarsitano, M. A., de Castro, F. A. B., Fernandes Júnior, F., et al. (2014). Performance, carcass and meat quality of ewes supplemented with magnesium oxide. *Rev. Bras. Zootec.* 43, 27–35. doi: 10.1590/S1516-35982014000100005

Cunningham, B. A., Moncur, J. T., Huntington, J. T., and Kinlaw, W. B. (1998). "Spot 14" protein: a metabolic integrator in normal and neoplastic cells. *Thyroid* 8, 815–825. doi: 10.1089/105072502761016494

Degerman, E., Ahmad, F., Chung, Y. W., Guirguis, E., Omar, B., Stenson, L., et al. (2011). From PDE3B to the regulation of energy homeostasis. *Curr. Opin. Pharmacol.* 11, 676–682. doi: 10.1016/j.coph.2011.09.015

Diniz, W. J., Coutinho, L. L., Tizioto, P. C., Cesar, A. S. M., Gromboni, C. F., Nogueira, A. R. A., et al. (2016). Iron content affects lipogenic gene expression in the muscle of nelore beef cattle. *PLoS One* 11:e0161160. doi: 10.1371/journal.pone.0161160

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Du, M., Shen, Q. W., Underwood, K. R., Tong, J. F., and Zhu, M. (2009). "AMP-activated protein kinase in muscle growth, fat deposition, and meat quality," in *Applied Muscle Biology and Meat Science*, eds M. Du and R. J. McCormick (Boca Raton, FL: CRC Press), 360.

Engle, T. E., Spears, J. W., Armstrong, T. A., Wright, C. L., and Odle, J. (2000). Effects of dietary copper source and concentration on carcass characteristics and lipid and cholesterol metabolism in growing and finishing steers. *J. Anim. Sci.* 78:1053. doi: 10.2527/2000.7841053x

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354

Geistlinger, L., da Silva, V. H., Cesar, A. S. M., Tizioto, P. C., Waldron, L., Zimmer, R., et al. (2018). Widespread modulation of gene expression by copy number variation in skeletal muscle. *Sci. Rep.* 8:1399. doi: 10.1038/s41598-018-19782-4

Gonçalves, T. M., de Almeida Regitano, L. C., Koltes, J. E., Cesar, A. S. M., da Silva Andrade, S. C., Mourão, G. B., et al. (2018). Gene Co-expression analysis indicates potential pathways and regulators of beef tenderness in nellore cattle. *Front. Genet.* 9:441. doi: 10.3389/fgene.2018.00441

Graugnard, D. E., Piantoni, P., Bionaz, M., Berger, L. L., Faulkner, D. B., and Loor, J. J. (2009). Adipogenic and energy metabolism gene networks in longissimus lumborum during rapid post-weaning growth in Angus and Angus x Simmental cattle fed high-starch or low-starch diets. *BMC Genomics* 10:142. doi: 10.1186/1471-2164-10-142

Hardie, D. G. (2011). AMP-activated protein kinase–an energy sensor that regulates all aspects of cell function. *Genes Dev.* 25, 1895–1908. doi: 10.1101/gad.17420111

Hay, S. M., Mcardle, H. J., Hayes, H. E., Stevens, V. J., and Rees, W. D. (2016). The effect of iron deficiency on the temporal changes in the expression of genes associated with fat metabolism in the pregnant rat. *Physiol. Rep.* 4, 1–9. doi: 10.14814/phy2.12908

Hoffman, N. J., Penque, B. A., Habegger, K. M., Sealls, W., Tackett, L., and Elmendorf, J. S. (2015). Chromium enhances insulin responsiveness via AMPK. *J. Nutr. Biochem.* 25, 565–572. doi: 10.1016/j.jnutbio.2014.01.007.Chromium

Je, W., Kim, M., Park, H., Sik, H., and Jae, M. (2006). AMPK activation increases fatty acid oxidation in skeletal muscle by activating PPARalpha and PGC-1. *Biochem. Biophys. Res. Commun.* 340, 291–295. doi: 10.1016/j.bbrc.2005.12.011

Kneeskern, S. G., Dilger, A. C., Loerch, S. C., Shike, D. W., and Felix, T. L. (2016). Effects of chromium supplementation to feedlot steers on growth performance, insulin sensitivity, and carcass characteristics. *J. Anim. Sci.* 94, 217–226. doi: 10.2527/jas.20159517

Koohmaraie, M., Kent, M. P., Shackelford, S. D., Veiseth, E., and Wheeler, T. L. (2002). Meat tenderness and muscle growth: Is there any relationship? *Meat Sci.* 62, 345–352. doi: 10.1016/S0309-1740(02)00127-4

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* 1:54. doi: 10.1186/1752-0509-1-54

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* 24, 719–720. doi: 10.1093/bioinformatics/btm563

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J. J., Chen, S., et al. (2018). Application of weighted gene co-expression network analysis for data from paired design. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-017-18705-z

Liu, Y., Du, M., Li, X., Chen, L., Shen, Q., Tian, J., et al. (2016). Role of the ubiquitin-proteasome pathway on proteolytic activity in postmortem proteolysis and tenderisation of sheep skeletal muscle. *Int. J. Food Sci. Technol.* 51, 2353–2359. doi: 10.1111/ijfs.13214

Loor, J. J. (2010). "Transcriptomics of muscle and fat development in growing cattle," in *Energy and Protein Metabolism and Nutrition*, ed. G. M. Crovetto (Italy: Wageningen Academic Publishers), 59–68.

Mateescu, R. (2014). "It is possible to genetically change the nutrient profile of beef," in *Proceedings of the Beef improvement Federation Research Symposium & Annual Meeting*, ed. Beef Improvement Federation (Lincoln, NE: Oklahoma State University), 87–92.

Mateescu, R. G., Garrick, D. J., and Reecy, J. M. (2017). Network analysis reveals putative genes affecting meat quality in Angus cattle. *Front. Genet* 8:171. doi: 10.3389/fgene.2017.00171

McCormick, R. J. (2009). "Collagen," in *Applied Muscle Biology and Meat Science*, eds M. Du and R. J. McCormick (Boca Raton, FL: CRC Press), 360.

Menendez, J. A., and Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* 7, 763–777. doi: 10.1038/nrc2222

Mihaylova, M. M., and Shaw, R. J. (2011). The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat. Cell Biol.* 13, 1016–1023. doi: 10.1038/ncb2329

Najafpanah, M. J., Sadeghi, M., Zali, A., Moradi-shahrebabak, H., and Mousapour, H. (2014). Chromium downregulates the expression of Acetyl CoA Carboxylase 1 gene in lipogenic tissues of domestic goats: a potential strategy for meat quality improvement. *Gene* 543, 253–258. doi: 10.1016/j.gene.2014.04.006

NCBI (2018). *ARS-UCD1.2*. Available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Bos_taurus/106/

Nishimura, T. (2015). Role of extracellular matrix in development of skeletal muscle and postmortem aging of meat. *Meat Sci.* 109, 48–55. doi: 10.1016/j.meatsci.2015.05.015

Oh, D., Lee, Y.-S., La, B., Lee, J.-Y., Park, Y., Lee, J., et al. (2014). Identification of exonic nucleotide variants of the thyroid hormone responsive protein gene associated with carcass traits and Fatty Acid composition in korean cattle. *Asian Australas. J. Anim. Sci.* 27, 1373–1380. doi: 10.5713/ajas.2014.14101

Omar, B., Zmuda-Trzebiatowska, E., Manganiello, V., Göransson, O., and Degerman, E. (2009). Regulation of AMP-activated protein kinase by cAMP in adipocytes: roles for phosphodiesterases, protein kinase B, protein kinase A. Epac and lipolysis. *Cell. Signal.* 21, 760–766. doi: 10.1016/j.cellsig.2009.01.015

Pamei, G., Gnanaraj, P., Sivakumar, T., Jawahar, K., Muthuramalingam, T., and Pothiappan, P. (2014). Influence of dietary supplementation of chromium on the carcass traits of crossbred pigs. *J. Adv. Vet. Anim. Res.* 1, 125–129. doi: 10.5455/javar.2014.a22

Pannier, L., Pethick, D. W., Boyce, M. D., Ball, A. J., Jacob, R. H., and Gardner, G. E. (2014). Associations of genetic and non-genetic factors with concentrations of iron and zinc in the longissimus muscle of lamb. *Meat Sci.* 96, 1111–1119. doi: 10.1016/j.meatsci.2013.08.013

Pighin, D., Pazos, A., Chamorro, V., Paschetta, F., Cunzolo, S., Godoy, F., et al. (2016). A contribution of beef to human health: a review of the role of the animal production systems. *Sci. World J.* 2016: 8681491. doi: 10.1155/2016/8681491

Pilkis, S. J., and Granner, D. K. (1992). Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis. *Annu. Rev. Physiol.* 54, 885–909. doi: 10.1146/annurev.ph.54.030192.004321

Ponsuksili, S., Du, Y., Hadlich, F., Siengdee, P., Murani, E., Schwerin, M., et al. (2013). Correlated mRNAs and miRNAs from co-expression and regulatory networks affect porcine muscle and finally meat properties. *BMC Genomics* 14:533. doi: 10.1186/1471-2164-14-533

Reverter, A., Hudson, N. J., Wang, Y., Tan, S.-H., Barris, W., Byrne, K. A., et al. (2006). A gene coexpression network for bovine skeletal muscle inferred from microarray data. *Physiol. Genomics* 28, 76–83. doi: 10.1152/physiolgenomics.00105.2006

Ritchie, H., and Roser, M. (2018). *Micronutrient Deficiency*. Available at: https://ourworldindata.org/micronutrient-deficiency [accessed August 22, 2018].

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Ropka-molik, K., Knapik, J., Pieszka, M., and Szmatoła, T. (2017). Nutritional modification of SCD, ACACA and LPL gene expressions in different ovine tissues. *Arch. Anim. Breed.* 60, 243–250. doi: 10.5194/aab-60-243-2017

Sadeghi, M., Javad, M., Panah, N., Reza, M., and Emami, A. (2015). Transcription analysis of genes involved in lipid metabolism reveals the role of chromium in reducing body fat in animal models. *J. Trace Elem. Med. Biol.* 32, 45–51. doi: 10.1016/j.jtemb.2015.05.004

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303.metabolite

Silva, V. H., Regitano, L. C., Geistlinger, L., Pértille, F., Giachetto, P. F., Brassaloti, R. A., et al. (2016). Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. *PLoS One* 11:e0157711. doi: 10.1371/journal.pone.0157711

Silva-Vignato, B., Coutinho, L. L., Cesar, A. S. M., Poleti, M. D., Regitano, L. C. A., and Balieiro, J. C. C. (2017). Comparative muscle transcriptome associated with carcass traits of Nellore cattle. *BMC Genomics* 18:506. doi: 10.1186/s12864-017-3897-x

Smith, L. R., Meyer, G., and Lieber, R. L. (2013). Systems analysis of biological networks in skeletal muscle function. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 5, 55–71. doi: 10.1002/wsbm.1197

Staniek, H., and Wójciak, R. W. (2018). The combined effect of supplementary Cr(III) propionate complex and iron deficiency on the chromium and iron status in female rats. *J. Trace Elem. Med. Biol.* 45, 142–149. doi: 10.1016/j.jtemb.2017.10.010

Suravajhala, P., Kogelman, L. J. A., and Kadarmideen, H. N. (2016). Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. *Genet. Sel. Evol.* 48:38. doi: 10.1186/s12711-016-0217-x

Tajima, T., Iijima, K., Watanabe, T., and Yamaguchi, H. (1981). The influence of calcium ions on the synthesis of collagen and glycosaminoglycans in human diploid cells in culture. *Exp. Pathol.* 19, 219–225. doi: 10.1016/S0232-1513(81)80067-9

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A., Di Nueda, M. J., Ferrer, A., et al. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* 43:e140. doi: 10.1093/nar/gkv711

Tizioto, P. C., Decker, J. E., Taylor, J. F., Schnabel, R. D., Mudadu, M. A., Silva, F. L., et al. (2013). Genome scan for meat quality traits in Nelore beef cattle. *Physiol. Genomics* 45, 1012–1020. doi: 10.1152/physiolgenomics.00066.2013

Tizioto, P. C., Gromboni, C. F., Nogueira, A. R., de Souza, M. M., Mudadu, M. D. A., Tholon, P., et al. (2014). Calcium and potassium content in beef: influences on tenderness and associations with molecular markers in Nellore cattle. *Meat Sci.* 96, 436–440. doi: 10.1016/j.meatsci.2013.08.001

Tizioto, P. C., Taylor, J. F., Decker, J. E., Gromboni, C. F., Mudadu, M. A., Schnabel, R. D., et al. (2015). Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle. *Genet. Sel. Evol.* 47:15. doi: 10.1186/s12711-014-0083-3

Verhagen, A. M., Coulson, E. J., and Vaux, D. L. (2001). Inhibitor of apoptosis proteins and their relatives: IAPs and other BIRPs. *Genome Biol.* 2, reviews3009.1–reviews3009.10. doi: 10.1186/gb-2001-2-7-reviews3009

Wang, L., and Michoel, T. (2016). "Detection of Regulator Genes and eQTLs in Gene Networks," in *Systems Biology in Animal Production and Health*, Vol. 1, ed. H. Kadarmideen (Cham: Springer International Publishing), 1–23. doi: 10.1007/978-3-319-43335-6_1

Xu, J., Ji, J., and Yan, X. (2012). Cross-Talk between AMPK and mTOR in regulating energy balance. *Crit. Rev. Food Sci. Nutr.* 52, 373–381. doi: 10.1080/10408398.2010.500245

Zhbannikov, I. Y., Hunter, S. S., Foster, J. A., and Settles, M. L. (2017). "SeqyClean: a pipeline for high-throughput sequence data preprocessing," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics ACM-BCB 17*, (New York, NY: ACM), 407–416. doi: 10.1145/3107411.3107446

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Diniz, Mazzoni, Coutinho, Banerjee, Geistlinger, Cesar, Bertolini, Afonso, de Oliveira, Tizioto, Kadarmideen and Regitano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 March 2019
doi: 10.3389/fgene.2019.00230

Systems Biology Reveals *NR2F6* and *TGFB1* as Key Regulators of Feed Efficiency in Beef Cattle

*Pâmela A. Alexandre[1,2], Marina Naval-Sanchez[2], Laercio R. Porto-Neto[2], José Bento S. Ferraz[1], Antonio Reverter[2] and Heidge Fukumasu[1]**

[1] Department of Veterinary Medicine, College of Animal Sciences and Food Engineering, University of São Paulo, Pirassununga, Brazil, [2] Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia

Systems biology approaches are used as strategy to uncover tissue-specific perturbations and regulatory genes related to complex phenotypes. We applied this approach to study feed efficiency (FE) in beef cattle, an important trait both economically and environmentally. Poly-A selected RNA of five tissues (adrenal gland, hypothalamus, liver, skeletal muscle and pituitary) of eighteen young bulls, selected for high and low FE, were sequenced (Illumina HiSeq 2500, 100 bp, pared-end). From the 17,354 expressed genes considering all tissues, 1,335 were prioritized by five selection categories (differentially expressed, harboring SNPs associated with FE, tissue-specific, secreted in plasma and key regulators) and used for network construction. *NR2F6* and *TGFB1* were identified and validated by motif discovery as key regulators of hepatic inflammatory response and muscle tissue development, respectively, two biological processes demonstrated to be associated with FE. Moreover, we indicated potential biomarkers of FE, which are related to hormonal control of metabolism and sexual maturity. By using robust methodologies and validation strategies, we confirmed the main biological processes related to FE in *Bos indicus* and indicated candidate genes as regulators or biomarkers of superior animals.

Keywords: feed efficiency, residual feed intake, Nellore (Zebu), *Bos indicus*, inflammation, muscle development, motif discovery, regulatory gene network

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Elisabetta Giuffra, INRA Centre Jouy-en-Josas, France
Carolina Neves Correia, University College Dublin, Ireland

***Correspondence:**
Heidge Fukumasu
fukumasu@usp.br

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *09 August 2018*
Accepted: *04 March 2019*
Published: *22 March 2019*

Citation:
Alexandre PA, Naval-Sanchez M, Porto-Neto LR, Ferraz JBS, Reverter A and Fukumasu H (2019) Systems Biology Reveals NR2F6 and TGFB1 as Key Regulators of Feed Efficiency in Beef Cattle. Front. Genet. 10:230. doi: 10.3389/fgene.2019.00230

INTRODUCTION

Since the domestication of the first species, animal selection aims to meet human needs and their changes over time. The current main selection goals in livestock production are increase of productivity, reduction of the environmental impact and reduction of competition for grains for human nutrition (Hayes et al., 2013). Thus, feed efficiency (FE) has become a relevant trait of study, as animals considered of high feed efficiency are those presenting reduced feed intake and lower production of methane and manure without compromising animal's weight gain (Gerber et al., 2013). However, the incorporation of FE as selection criteria in animal breeding programs is costly and time consuming. Daily feed intake and weight gain for a large number of animals need to be recorded for at least 70 days to obtain accurate estimates of FE (Archer et al., 1997).

In the past years, several studies have been carried out with the aim to identify molecular markers associated with FE to enable a faster and cost-effective identification of superior animals

(Rolf et al., 2011; Oliveira et al., 2014; Santana et al., 2014; Seabury et al., 2017). However, for each population, different biological processes seem to be identified (Rolf et al., 2011; Oliveira et al., 2014; Santana et al., 2014; Seabury et al., 2017). Probably, that is because FE is a multifactorial trait and many different biological mechanisms seems to be involved in its regulation (Herd et al., 2004; Herd and Arthur, 2009). It has been demonstrated that high FE animals present increased mitochondrial function (Connor et al., 2010; Lancaster et al., 2014), less oxygen consumption (Gonano et al., 2014) and delayed puberty (Shaffer et al., 2011; Randel and Welsh, 2013; Fontoura et al., 2016). On the other hand, low FE animals have increased physical activity, ingestion frequency and stress level (Kelly et al., 2010; Cafe et al., 2011; Chen et al., 2014; Francisco et al., 2015), increased leptin and cholesterol levels (Nkrumah et al., 2007; Alexandre et al., 2015; Foote et al., 2016; Mota et al., 2017), higher subcutaneous and visceral fat (Mader et al., 2009; Gomes et al., 2012; Santana et al., 2012), higher energy wastage as heat (Archer et al., 1999; Montanholi et al., 2009, 2010) and more hepatic lesions associated with inflammatory response (Alexandre et al., 2015; Paradis et al., 2015).

In the context of such a complex trait, we perform a multiple-tissue transcriptomic analysis of high (HFE) and low (LFE) feed efficient Nellore cattle across tissues related to endocrine control of hunger/satiety, hydric and energy homeostasis, stress and immune response, physical and sexual activity, as is the case of hypothalamus-pituitary-adrenal axis and organs as liver and skeletal muscle. Using gene co-expression across tissues and conditions, we derived a regulatory network revealing NR2F6 and TGFB1 signaling as key regulators of hepatic inflammatory response and muscle tissue development, respectively. Next, we applied advanced motif discovery methods which (i) validate that co-expressed genes are enriched for NR2F6 and TGFB1 signaling effector molecule SMAD3 binding sites in their 10 kb upstream regions and (ii) predict direct transcription factor (TF) – Target gene (TG) interactions at the sequence level. These binding interactions were experimentally validated with public TF ChIP-seq from ENCODE (Encode Project Consortium, 2012; Sloan et al., 2016). Regulatory activity in the tissues of interest was also confirmed by performing an enrichment analysis on open chromatin tracks and histone chromatin marks across cell types and tissues in the human and cow genome. Moreover, we propose a hormonal control of differences in metabolism and sexual maturity between HFE and LFE animals, indicating potential biomarkers for further validation such as adrenomedullin, FSH, oxytocin, somatostatin and TSH.

RESULTS

Multi-Tissue Transcriptomic Data Reveal Differences Between High and Low Feed Efficient Animals

Feed efficiency is a complex trait characterized by multiple distinct biological processes including metabolism, ingestion, digestion, physical activity and thermoregulation (Herd et al., 2004; Herd and Arthur, 2009). To study FE at transcriptional level we performed RNAseq of five tissues (i.e., adrenal gland, hypothalamus, liver, muscle and pituitary) from nine male bovines of high feed efficiency [HFE, characterized by low residual feed intake (RFI) (Koch et al., 1963)] and nine of low FE (LFE, characterized by high RFI). In total, we analyzed 18 samples of liver, hypothalamus and pituitary; 17 of muscle and 15 of adrenal gland, yielding 13 million reads per sample on average (**Supplementary Table 1**). Gene expression was estimated for 24,616 genes present in the reference genome (UMD 3.1) and after quality control (refer to Section "Materials and Methods"), 17,354 genes were identified as being expressed in at least one of the five tissues analyzed.

Differential expression (DE) analysis between HFE and LFE animals resulted in 471 DE genes across tissues ($P < 0.001$, **Supplementary Image 1**), namely, 111 in adrenal gland, 125 in hypothalamus, 91 in liver, 104 in muscle and 98 in pituitary (**Supplementary Tables 2A–E**). Although no significant functional enrichment was found for the 281 genes up-regulated in HFE group, the 248 genes down-regulated presented a significant enrichment of GO terms such as response to hormone (Padj = 5.43×10^{-6}), regulation of hormone levels (Padj = 3.48×10^{-6}), cell communication (Padj = 3.18×10^{-4}), regulation of signaling receptor activity (Padj = 3.20×10^{-4}), hormone metabolic process (Padj = 5.86×10^{-4}), response to corticosteroid (Padj = 6.28×10^{-4}), regulation of secretion (Padj = 7.2×10^{-4}), response to lipopolysaccharide (Padj = 7.9×10^{-4}) and regulation of cell proliferation (Padj = 1.86×10^{-3}). Refer to **Supplementary Image 2** to see all enriched terms.

Overlap Between Gene Selection Criteria Prioritizes Genes Associated With Feed Efficiency

The genetic architecture behind complex traits involves a large variety of genes with coordinated expression patterns, which can be represented by gene regulatory networks as a blueprint to study their relationships and to identify central regulatory genes (Swami, 2009). Therefore, it is important to select relevant genes and gene families according to the phenotype of interest to be used for network analysis. We defined five categories of genes (see Section "Materials and Methods" for further information) for inclusion in co-expression analysis: (1) differentially expressed (DE), (2) genes harboring SNPs previously associated with FE (harboring SNP), (3) tissue specific (TS), (4) genes coding proteins secreted in plasma by any of the five tissues analyzed (secreted) and (5) key regulators.

As reported before, we have identified 471 DE genes between HFE and LFE animals (**Figure 1A** and **Supplementary Table 3A**). In addition, 267 genes were selected for harboring SNPs previously associated with FE, as not only differences in expression levels can influence the phenotype but also polymorphism in the DNA sequence that can alter the translated protein behavior (**Supplementary Table 3B**). Moreover, 396 were selected for being tissue specific (refer to Section "Materials and Methods" for definition); 22 in adrenal gland,

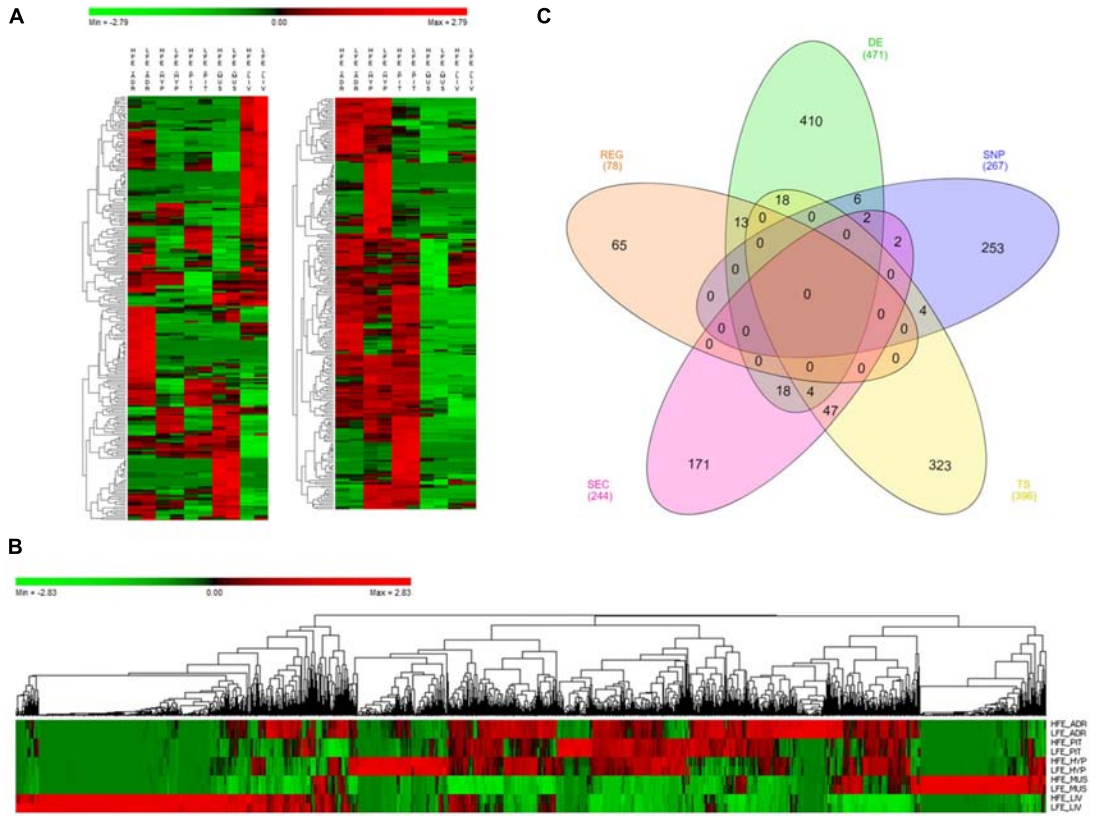


FIGURE 1 | Genes selected for co-expression network construction. **(A)** Heatmap of normalized mean expression (NME) of 471 differentially expressed (DE) genes between high (HFE) and low (LFE) feed efficient animals in adrenal gland (ADR), hypothalamus (HYP), liver (LIV), muscle (MUS) and pituitary (PIT). Genes (rows) and samples (columns) are organized by hierarchical clustering based on Euclidean distances. **(B)** NME heatmap of all 1,335 genes selected for network construction. Genes (columns) and samples (rows) are organized by hierarchical clustering based on Euclidean distances. **(C)** Venn diagram of 1,335 genes selected for network construction. The inclusion criteria for selecting genes were divided into five categories: differentially expressed genes (DE), tissue specific genes (TS), genes harboring SNPs reported in literature as being associated with feed efficiency in beef cattle (SNP), genes encoding proteins secreted by at least one of the tissues in plasma (SEC) and key regulators (REG). Numbers between brackets indicate the total number of genes in each category.

32 in hypothalamus, 215 in liver, 118 in muscle and 9 in pituitary (**Supplementary Table 3C**). A total of 244 genes coding proteins secreted in plasma were selected because of their potential as biomarkers of FE (**Supplementary Table 3D**). From those, 135 had liver as the tissue of maximum expression and were functionally enriched for GO terms such as complement activation (Padj = 1.82×10^{-19}), regulation of acute inflammatory response (Padj = 1.89×10^{-14}), innate immune response (Padj = 9.71×10^{-12}), negative regulation of endopeptidase activity (Padj = 2.35×10^{-10}), platelet degranulation (Padj = 1.08×10^{-10}), regulation of coagulation (Padj = 3.39×10^{-9}), triglyceride homeostasis (Padj = 1.23×10^{-6}), cholesterol efflux (Padj = 1.03×10^{-5}) (**Supplementary Image 3**). Finally, from 1570 potential regulators in publicly available Animal TFdb, 78 were identified as key regulators of the genes selected by all the other categories, i.e., 78 genes presented a coordinated expression level with many of the genes in the network reflecting a tight control of expression patterns across tissues (**Supplementary Table 3E**).

Considering all the inclusion criteria, 1,335 genes were selected to be included in co-expression network analysis (**Figure 1B** and **Supplementary Table 4**), some of them selected in more than one category (**Figure 1C**). Regarding DE genes, six of them were also reported before as harboring SNPs associated with the phenotype (*LUZP2, MAOB, SFRS5, SLC24A2, SOCS3* and *WIF1*) (Bolormaa et al., 2011; Saatchi et al., 2014; Ramayo-Caldas et al., 2018) and 13 of them were key regulators (*HOPX, PITX1, CRYM, PLCD1, ND6, CYTB, ND1, MT-ND4L, ND5, ATP8, ND4, ENSBTAG00000046711* and *ENSBTAG00000048135*). Many of the genes that are both DE and regulators are involved in respiratory chain (*ND6, CYTB, ND1, MT-ND4L, ND5, ATP8* and *ND4*) and were all up-regulated in HFE group.

Considering both DE and secreted genes, 18 were identified (*NOV, SPP1, CTGF, OXT, PTX3, VGF, CCL21, COL1A2, PGF, SOD3, SERPINE1, PRL, PON1, SST, JCHAIN, PCOLCE, IGFBP6* and *SCG2*). In addition, four genes were DE, secreted and tissue specific, two from liver (*CXCL3* and *IGFBP1*) and two from pituitary (*NPY* and *CYP17A1*). Genes *RARRES2* and *PENK* (proenkephalin) were DE, secreted and had been previously reported as harboring SNP associated with FE (RARRES2:AnimalQTLdb Release 35 – QTL:20671,

rs133399845; PENK: Bolormaa et al. (2011)- rs136198266, rs134428213, rs137492938, rs132881564). Other DE genes worthy to highlight, due to their well-known role in metabolic processes, are *AMH* (anti-mullerian hormone), *TSHB* (thyroid stimulating hormone beta), *FGF21* (Fibroblast growth factor 21) and *FST* (follistatin), up-regulated in HFE group, and *PMCH* (pro-melanin concentrating hormone), *ADM* (adrenomedullin) and *FSHB* (follicle stimulating hormone beta), up-regulated in LFE group.

Co-expression Network Reveals Regulatory Genes and Biological Processes Related to Feed Efficiency

The co-expression network (**Figure 2**) was composed of 1,317 significant genes and 91,932 connections, with a mean of 70 connections per gene (considering only genes with significant expression correlation $\geq |0.90|$). Most of the connections (51%) involved a DE gene and 23% of those were between two DE genes. Tissue specific (TS) genes were involved in 49% of the connections with 119 connections per gene on average, which was higher than the overall network mean and reflects the close relationship among genes involved in tissue specific functions. Key regulators were the least represented category in the network (only 78 genes) but accounted for 11% of the connections in the network with the highest value of mean connections per gene, 131 connections, which is in accordance with their regulatory role. Regarding the connections within tissues, when we ranked all the genes in the network by the number of connections and looked at the top 50 genes, 29 were from liver, 15 were from muscle and 3, 2 and 1 were from pituitary, adrenal gland and hypothalamus, respectively. These results indicate very well-coordinated expression patterns in liver and muscle that could be a reflection of the number of TS genes in those tissues and the presence of central regulatory genes coordinating the expression of many other genes.

In the network (**Figure 2**), genes were grouped together by tissue which was mostly driven by TS genes. As mentioned before, most of the secreted protein-coding genes were located in the liver. Most of the key regulators were located peripherally in relation to the clusters which could be reflecting their regulatory nature independent of tissue specificity. Despite that, some regulators draw attention because of their high number of connections.

The top five most connected regulators were *EPC1, NR2F6, MED21, ENSBTAG00000031687* and *CTBP1*, varying from 317 to 284 connections. They were all first neighbors of each other and were connected mainly to genes with higher expression in liver and essentially enriched for acute inflammatory response (Padj = 4.5×10^{-13}, **Supplementary Image 4**). The next most connected regulator is *TGFB1* with 217 connections. It is mainly connected to genes from muscle that are primarily enriched for muscle organ development (Padj = 6.87×10^{-5}) and striated muscle contraction (Padj = 1.39×10^{-5}, **Supplementary Image 5**). Besides indicating main regulator genes, the gene co-expression networks approach can be useful to access the role of specific genes. For instance, gene *FGF21*, a hormone up-regulated in liver of HFE animals, is directly connected to genes enriched for plasma lipoprotein particle remodeling, regulation of lipoprotein oxidation and cholesterol efflux (Padj = 5.64×10^{-3}, **Supplementary Image 6**). Indeed, according to the literature, this gene is associated with decrease in body weight, blood triglycerides and LDL-cholesterol (Cheung and Deng, 2014).

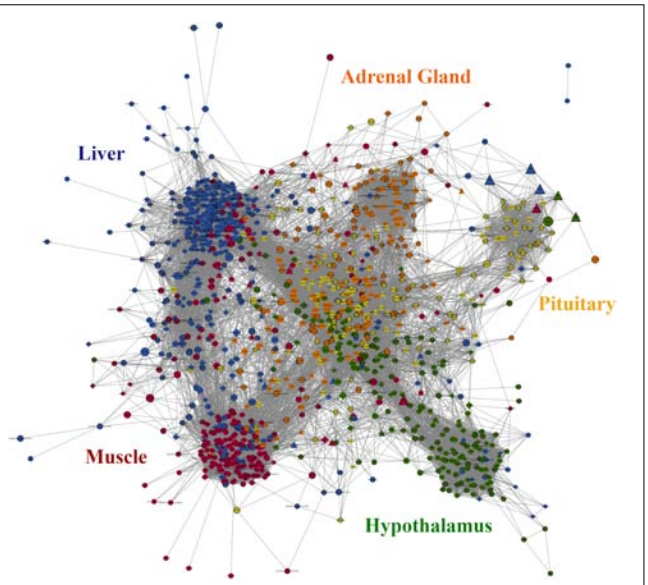


FIGURE 2 | Gene co-expression network constructed using PCIT algorithm on 1,335 selected genes (see Section "Materials and Methods"). Only significant correlations above | 0.9| and their respective genes were considered, totaling to 1,317 genes and 91,932 connections. Nodes with diamond shape correspond to genes coding for proteins secreted in plasma and triangles correspond to key regulators; all the other genes are represented by ellipses. Nodes with black borders are differentially expressed between high and low feed efficiency groups. Colors are relative to the tissue of maximum expression: blue represents liver, red represents muscle, yellow represents pituitary, green represents hypothalamus and orange represents adrenal gland. The size of the nodes is relative to the normalized mean expression values in all samples.

Motif Discovery Confirms NR2F6 as a Key Regulator of Liver Transcriptional Changes Between High and Low Feed Efficiency

By means of the power-law theory, co-expression networks present many nodes with few connections and few central nodes with many connections (de la Fuente, 2010), being the last ones indicated as central regulatory genes responsible for the transcriptional changes between the divergent phenotypes analyzed. In our study, the most connected regulators were indicated, together with their target genes, i.e., their first neighbors in the network. Those genes are a mixture of direct and indirect regulator targets. In order to validate the regulatory role of the most connected regulators in the network and identify their core direct targets, we performed motif discovery in their co-expressed target genes. It is noteworthy that motif discovery should confirm the presence of DNA motifs of a

TF in the regulatory regions of co-expressed genes. From the top five most connected regulators from our previous co-expression analysis, only NR2F6 has the ability to bind DNA. In contrast, the other four regulators act mainly as cofactors (corepressor, i.e., CTBP1; coactivator, i.e., MED21; or histones modifier, i.e., EPC1), that is co-binding through protein–protein interactions.

The analysis of 313 co-expressed genes with *NR2F6* (**Figure 3A**) yield the Nuclear Factor motif HNF4-NR2F2 (transfac_pro-M01031) as the second motif most enriched out of 9732 PWMs (position weight matrices) with a Normalized Enrichment Score (NES) of 7.98 (**Figure 3B**). In addition, a total of 19 motifs associated with HNF4-NR2F2 were enriched in the dataset, associating HNF4-NR2F2 to 168 direct target genes (**Figure 3C**). Due to motif redundancy or highly similarity between a plethora of TFs, these motifs can be associated with multiple TFs from HNF4 (direct) to several nuclear factors such as NR2F6 (motif similarity score FDR 1.414×10^{-5}). However, our co-expression analysis strongly indicates that NR2F6 is the key TF, since it was the TF with the highest number of nodes in the co-expression network (**Figure 3C**) and neither HNF4 nor NR2F2 were prioritized by any selection category to be included in the network.

Each of the NR2F6 inferred direct target genes contain one or more predicted enhancers, i.e., regions with high-scoring motif binding sites for NR2F6 or TFs with highly similar motifs. To validate the binding of these genomic regions by NR2F6 or TFs with highly similar motifs to NR2F6, we performed a region enrichment analysis of our predicted NR2F6 binding sequences against public TF ChiP-seq bound regions in human cell lines from the ENCODE consortium (1394 TF binding site tracks, Encode Project Consortium, 2012; Sloan et al., 2016). This analysis confirms the experimental binding of TFs with similar binding as NR2F6 in HepG2 cells. In particular, HNF4A on human HepG2 (ENCFF001UGH, GSM803460, NES = 9.57), HNF4G (ENCFF001UGI, GSM803404, NES = 7.83), RXRA (ENCFF001UHJ, GSM803404, NES = 6.85), and NR2F2 (ENCFF001UGV, GSM1010810, NES = 4.45,) as the most enriched tracks (**Supplementary Data Sheet 1**). Recent NR2F6 ChIP-seq data on HepG2 (ENCODE experiment ENCSR518WPL, GSE96210) also confirms an enrichment for NR2F6, indicating predicted *NR2F6* binding regions are experimentally bound by NR2F6 in hepatocyte cell lines (**Figure 3D**).

Next, to validate that the NR2F6 binding in those regions is functional in liver we performed an enrichment analysis for open-chromatin (tracks = 655) and histone modifications (tracks = 2450) related to active regulatory elements (**Supplementary Table 5**). This analysis yielded DNA-seq on human hepatocytes (ENCFF001SOV, GSM816663, NES = 4.10), and H3K29ac and H3K4me3 in adult liver (Roadmap Epigenomics Consortium et al., 2015; GSM621630, GSM537709, respectively) as the most enriched tracks, respectively, strongly indicating that not only predicted target enhancers are bound by NR2F6 in Hepatocyte cell lines, but these regulatory regions are functionally active in hepatocytes and human liver (**Figure 3D**).

Regarding the cow genome, a recent open-chromatin study (Villar et al., 2015) has mapped active promoters and enhancers by H3K4me3 and H3K27ac ChIP-seq in cow liver resulting in 13,796 promoter and 45,786 enhancers. We performed an enrichment analysis of predicted *NR2F6* enhancers converted to cow coordinates (n = 779, **Supplementary Table 6**, Array Express Accession number E-MTAB-2633) resulting in 446 regions being identified as functional regulatory regions in cow liver. This number is significantly higher compared to the only 43 regions expected to overlap by random (1000 permutation tests) (**Figure 3E**).

Finally, in addition to *NR2F6* motif, *HNF1A* motif was found as a potential co-regulator in liver, in particular swissregulon-HNF1A.p2 with a NES = 10.17 and in total 20 enriched motifs and 170 direct targets were associated to *HNF1A* (**Figure 3B**). *HNF1* is a master regulator of liver gene expression (Tronche and Yaniv, 1992), thus making its finding justified.

Motif Discovery Validates TGFB1 Signaling Through SMAD3/MYOD1 Binding as Drivers of Transcriptional Differences in Muscle of Divergent Feed Efficient Cattle

The analysis of the 217 genes co-expressed with TGFB1 (**Figure 4A**) showed that most target genes motifs were enriched for master regulators of muscle differentiation, namely, MEF2 (NES = 10.42), a MADS box Transcription factor with 148 target genes, and MYOD1 (NES = 5.09), a bHLH transcription factor (CANNTG) with 136 direct target genes (**Figure 4B** and **Supplementary Data Sheet 2**). To evaluate the precision of our predicted MYOD1 (bHLH) target genes, we assessed how many of these TF-TG relationships had been previously experimentally reported. Based on MYOD1 ChIP-seq binding in mouse myotubules, 86 genes had already been associated with MYOD1 resulting in a 63% success rate (hypergeometric test 1.72×10^{-22}). SMAD3, the effector molecule of TGFB1 signaling, is known to recruit MYOD1 to drive transcriptional changes during muscle differentiation (Mullen et al., 2011). Thus, we evaluated whether predicted MYOD1 target genes were enriched for known SMAD3 target genes resulting in 21 out of 135 MYOD1 predicted target genes presenting SMAD3 ChIP-seq binding in myotubes, thus indicating that there is a statistically significant association between MYOD1 target genes and SMAD3 target genes in myotubes (hypergeometric test 1.98×10^{-6}) (**Figure 4C**) (Mullen et al., 2011). By contrast, no significant association was found between predicted MYOD1 target genes in this study and SMAD3 target genes in other cell lines, such as pro-B and ES cell (hypergeometric test 0.056 and 0.076, respectively) (Mullen et al., 2011). That is in agreement with the fact that the effect of TGFB1 signaling driven by SMAD3 DNA binding is tissue-specific (Liu et al., 2001). Our analysis predicted 621 potential MYOD1 binding sites, of which 114 (18%, **Supplementary Tables 7**, **8**) and 152 (24.5%, **Supplementary Table 9**) present a MYOD1 ChIP-seq signal in mouse C2C12 myotubes cells (Mullen et al., 2011) and in primary myotubes (Cao et al., 2009), respectively.

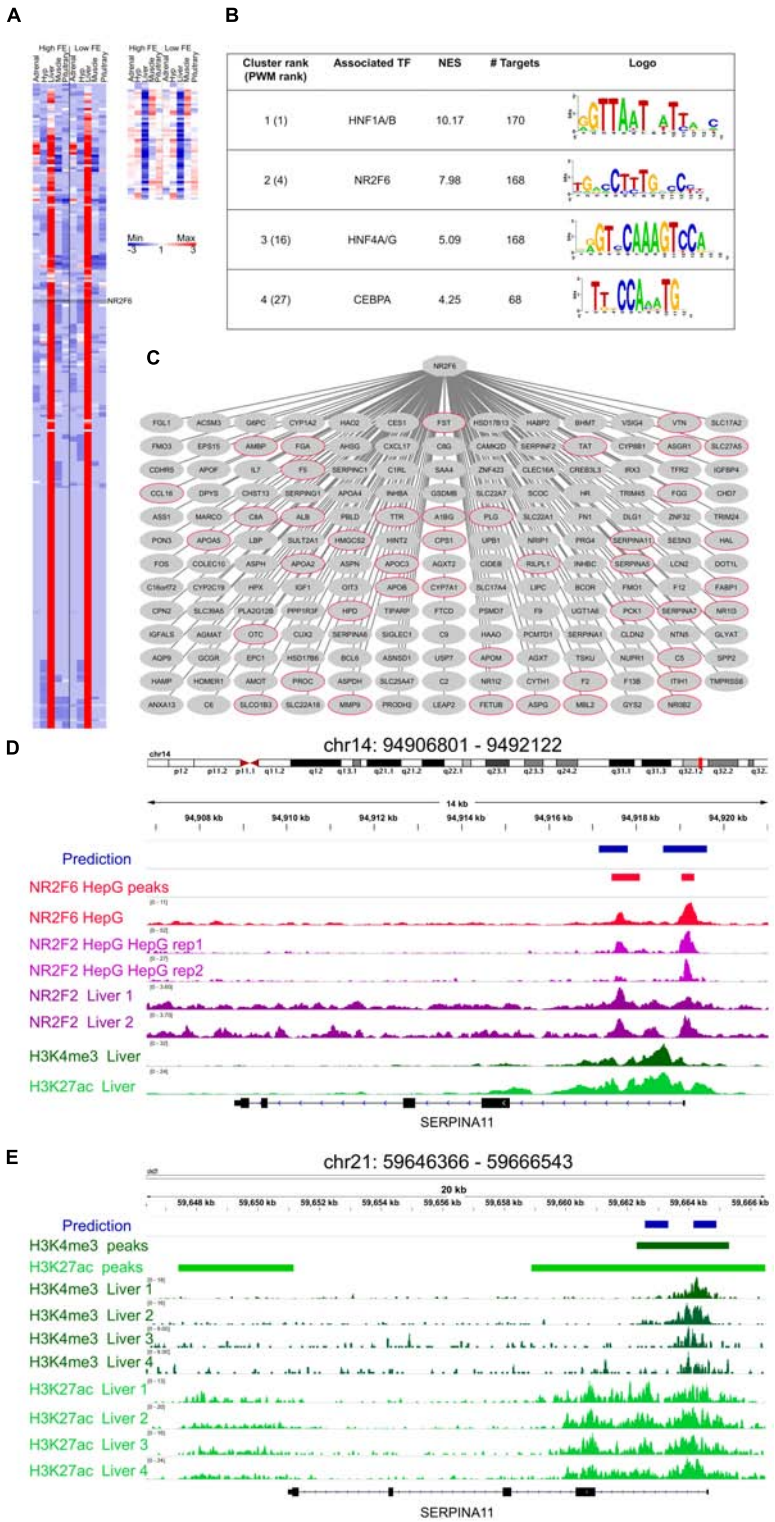


FIGURE 3 | Mapping of NR2F6 direct targets. **(A)** Heatmap of the 313 genes co-expressed with NR2F6 across all samples (derived from the co-expression analysis), **(B)** i-Regulon motif discovery results on the genes shown in panel **(A)**, **(C)** Predicted NR2F6 targetome. A red node indicates genes known to be targeted by NR2F6 in human Hepatocytes. **(D)** Example of predicted NR2F6 target regions for *SERPINA1* gene. The predicted enhancer overlaps the exact position for NR2F6 and NR2F2 binding in HepG sites from the ENCODE dataset as well as histone chromatin marks related to active regulatory regions, namely H3K27ac, and promoters, H3K4me3 in human primary tissue from RoadMap Epigenetics **(E)** The enhancer prediction in cow coordinates (bosTau6) overlaps a region marked with H3K4me3 in cow liver (Villar et al., 2015).

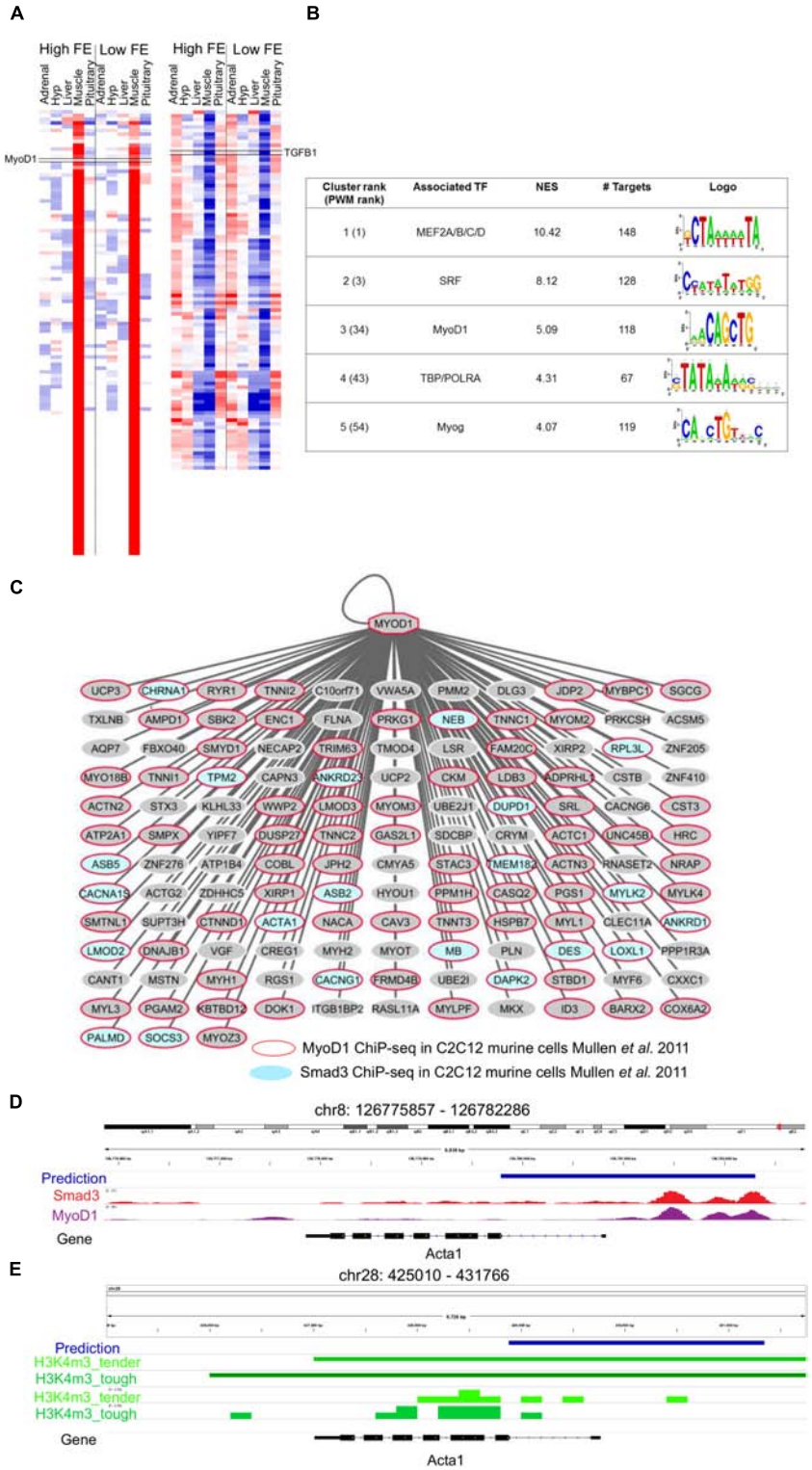


FIGURE 4 | Mapping the downstream network of TGFB1 signaling through SMAD3/MyoD1 DNA binding. **(A)** Heatmap of the 217 genes co-expressed with *TGFB1* (derived from the co-expression analysis). **(B)** i-Regulon motif discovery results on the genes shown in panel **(A)**, **(C)** Predicted MyoD targetome. A red node indicates genes know to be targeted my MyoD1 in murine myotubes (Mullen et al., 2011). Blue nodes indicate genes to be targeted by SMAD3, the effector DNA binding molecular of *TGFB1* signaling, in murine myotubes (Mullen et al., 2011). **(D)** Example of predicted MyoD1 target regions for *ACTA1* gene. The predicted enhancer overlaps the exact position for SMAD3 and MyoD1 ChIP-seq binding in murine myotubes (Mullen et al., 2011). **(E)** The enhancer prediction in cow coordinates (bosTau6) overlaps a promoter region marked with H3K4me3 in muscle tissue in cow (Zhao et al., 2015).

Finally, we evaluate whether predicted MYOD1 binding regions were regulatory regions active in muscle cells across different species, namely human, mouse and cow. To tackle this issue we performed an enrichment analysis across 2113 open-chromatin ENCODE tracks (Encode Project Consortium, 2012; Sloan et al., 2016). This analysis resulted in a clear enrichment of our predicted MYOD1 binding regions with H3K27ac (NES = 15.98) and H3K9ac (NES = 8.78) regions in the skeletal muscle (**Figure 4D**). Both chromatin marks are associated with active transcription, H3K27ac related to active enhancers and H3K9ac related to active gene transcription (Shin et al., 2012), thus validating most of our enhancer predictions that MYOD1 in human is active in the skeletal muscles. In cow, we assessed the overlap of predicted MYOD1 enhancers and promoter regions in cow muscle experimentally detected with H3K4me3 (Zhao et al., 2015). This resulted in 275 regions out of 653 (42%) overlap when only 11 regions are expected to overlap by random 1000 permutation test) (**Figure 4E**, **Supplementary Table 10**).

Differential Co-expression

Although the general co-expression network provides important insights about regulatory genes and their behavior, by creating specific networks for HFE and LFE and comparing the connectivity of the genes in each one, we can identify genes that change their behavior depending on the situation, moving from highly connected to lowly connected and vice-versa. We were able to identify 87 differentially connected genes between HFE and LFE ($P < 0.05$); 63 mainly expressed in liver, 19 in muscle and 3, 1 and 1 in hypothalamus, adrenal gland and pituitary, respectively (**Supplementary Table 11**). Those genes were enriched for terms such as regulation of blood coagulation (Padj = 3.14×10^{-10}), fibrinolysis (Padj = 7.71×10^{-7}), platelet degranulation (Padj = 7.49×10^{-6}), regulation of peptidase activity (Padj = 6.16×10^{-4}), antimicrobial humoral response (Padj = 2.49×10^{-3}), acute inflammatory response (Padj = 2.18×10^{-4}) and induction of bacterial agglutination (Padj = 3.58×10^{-2}) (**Supplementary Image 7**). It is important to highlight that 20 of the differentially connected genes were also differentially expressed (**Table 1**) and three of them, i.e., *SST, JCHAIN* and *IGFBP1*, were secreted in plasma as well, which make them very promising potential biomarkers.

DISCUSSION

Feed efficiency is a complex trait, regulated by several biological processes. Thus, the identification of genomic regions associated with this phenotype, as well as regulators genes and biomarkers to select superior animals and to direct management decisions, is still a great challenge. In this work, multi-tissue transcriptomic data of high and low feed efficient Nellore bulls were analyzed through robust co-expression network methodologies in order to uncover some of the biology that governs these traits and put forward candidate genes to be the focus of further research. In this sense, the validation of target genes of main transcription factors (key regulators) in the network by motif search proves the efficacy of the methodology for network construction and prioritizes some transcription factors as central regulators (Aerts et al., 2010;

TABLE 1 | Differentially connected and differentially expressed genes between high and low feed efficiency.

Gene name	Number of connections		Category*	Tissue of maximum expression	Tissue of differential expression
	Low feed efficiency	High feed efficiency			
SST	0	45	DE, SEC	Hypothalamus	Hypothalamus
SNORA73	41	108	DE	Liver	Liver
ENSBTAG00000047700	56	111	DE	Liver	Liver
ENSBTAG00000047121	62	111	DE	Liver	Liver
ENSBTAG00000047816	53	96	DE	Liver	Liver
ENSBTAG00000039928	50	89	DE	Liver	Liver
ANXA13	115	63	DE	Liver	Liver
FST	113	56	DE	Liver	Liver
PBLD	115	55	DE	Liver	Liver
ENSBTAG00000021368	95	0	DE	Liver	Liver
JCHAIN	52	113	DE, SEC	Liver	Liver
IGFBP1	55	0	DE, TS, SEC	Liver	Liver
SBK2	0	70	DE	Muscle	Muscle
ACTC1	54	0	DE	Muscle	Muscle
MYH1	0	47	DE, TS	Muscle	Muscle
HR	119	50	DE	Pituitary	Muscle
TAGLN	83	31	DE	Adrenal	Muscle, Pituitary
SFRP2	41	91	DE	Hypothalamus	Pituitary
FN1	119	69	DE	Liver	Pituitary
CAV1	98	50	DE	Muscle	Pituitary

**Differentially expressed genes between high and low feed efficiency (DE), tissue specific genes (TS) and genes encoding proteins secreted in plasma (SEC).*

Naval-Sańchez et al., 2013; Potier et al., 2014). Moreover, the addition of a category of genes coding proteins secreted in plasma in the co-expression analysis highlights the genes with potential to be explored as biomarkers of feed efficiency. We were able to identify genes related to main biological processes associated with feed efficiency and indicate key regulators.

Firstly, it is important to state that the 98 animals used to select the HFE and LFE groups in this study have been previously analyzed with regard to several phenotypic and molecular measures (Alexandre et al., 2015; Mota et al., 2017; Novais et al., 2019). It was observed that HFE and LFE groups had similar body weight gain, carcass yield and loin eye area but LFE animals had higher feed intake, greater fat deposition, higher serum cholesterol levels, as well as hepatic inflammatory response, indicated by transcriptome analysis of liver biopsy and proved by the higher number of periportal mononuclear infiltrate (histopathology) and increased serum gamma-glutamyl-transferase (GGT, a biomarker of liver injury) in this group (Alexandre et al., 2015). In the present study, the simultaneous analysis of five distinct tissues revealed the importance of hepatic tissue. Liver presented the most connected genes in the network, the largest number of differentially connected genes and the largest number of secreted genes, which, although can be explained by its biological function, are enriched mostly for terms related to lipid homeostasis and inflammatory response. Moreover, the top five most connected regulators in the network are co-expressed mainly with genes highly expressed in liver and also enriched for inflammatory response.

The relationship between FE and genes or pathways related to immune response and lipid metabolism is becoming more evident, as recent studies also reported in beef cattle (Karisa et al., 2014; Paradis et al., 2015; Weber et al., 2016; Zarek et al., 2017; Mukiibi et al., 2018) and pigs (Gondret et al., 2017; Ramayo-Caldas et al., 2018). In our previous work (Alexandre et al., 2015), we proposed that increased liver lesions associated with higher inflammatory response in the liver of LFE animals could be due to increased lipogenesis and/or higher bacterial infection in the liver. While further evidence is needed to test these hypotheses, the enrichment of terms such as induction of bacterial agglutination and response to lipopolysaccharide makes bacterial infection a strong possibility. Indeed, pigs with low FE were reported to have a increased risk of intestinal inflammation, higher neutrophil infiltration biomarkers and increased serum endotoxin (lipopolysaccharide and other bacterial products) which could be related to increased bacterial infection or to decreased capacity to neutralize endotoxins (Mani et al., 2013). The authors hypothesized that differences in bacterial population could partially explain the increase in circulating endotoxins, which could also be true for cattle given that differences in intestinal and ruminal bacterial population between high and low FE animals have already been reported (Myer et al., 2015, 2016). Furthermore, the literature reports lipopolysaccharides (LPS) may cause up-regulation of adrenomedullin (*ADM*) hormone (Shindo et al., 1998), an up-regulated gene in LFE individuals as showed here. It was also demonstrated in rats that intravenous infusion of LPS caused up-regulation of *ADM* in ileum, liver, lung, aorta, skeletal muscle and blood vessels (Shoji et al., 1995) whereas in our study, *ADM* presented differential expression in muscle, but not in liver.

Against pathogen invasion, a tightly regulated adaptive immune response must be triggered in order to allow T lymphocytes to produce cytokines or chemokines and B cells to differentiate and produce antibodies (Hermann-Kleiter and Baier, 2014). This regulation is known to be strongly influenced by the expression level and transcriptional activity of several nuclear receptors, including the NR2F-family, which consists of three orphan receptors: NR2F1, NR2F2 and NR2F6 (Hermann-Kleiter and Baier, 2014). Those receptors present highly conserved DNA and ligand binding domains among each other and across species (Pereira et al., 2000), and all three are expressed in adaptive and immune cells (Hermann-Kleiter and Baier, 2014). In our study, NR2F6 appeared as the second most connected regulator gene in the network while the other family members, although present in our expression data, were not selected by any of our inclusion criteria, thus indicating they might not be so relevant in our conditions. Indeed, NR2F6 appears to be a critical regulatory factor in the adaptive immune system by directly repressing the transcription of key cytokine genes in T effector cells (Hermann-Kleiter et al., 2008; Klepsch et al., 2016). The role of NR2F6 as a key regulator of inflammatory response in our network was validated at gene level by the identification of the binding motif HNF4-NR2F2 (transfac_pro-M01031) as one of the most enriched in NR2F6 target genes, due to the high similarity between NR2F2 and NR2F6 binding sites. Furthermore, using open chromatin data publicly available, we provided experimental evidence of the binding of TFs with highly similar binding motifs as NR2F6 in hepatocyte cells in humans and in cattle, thus, indicating that predicted target enhancers are functional in this tissue.

Another regulator prioritized in our analysis is *TGFB1*, the sixth most connected gene in the co-expression network, and a potential driver of transcriptional changes between high and low FE cattle in muscle. This gene has been previously described as a master regulator of FE in beef cattle, using genomics and metabolomics data (Widmann et al., 2015). Moreover, our motif discovery analysis showed that *TGFB1* co-expressed genes are mostly enriched for binding sites of master regulators of muscle differentiation such as MEF2 and MYOD. Indeed, public available data show many of TGFB1 target genes were associated with MYOD (Mullen et al., 2011). As it is known, signaling pathways are an effective mechanism for cells to respond to environmental cues by regulating gene expression. TGFB1 signaling triggers the phosphorylation of SMAD2/3 transcription factors, which co-bind with cell-type master regulators at the nuclear level allowing/triggering/leading to cell-type specific transcriptional changes (Schmierer and Hill, 2007; Mullen et al., 2011). In skeletal muscle cells, myoblasts and myotubes, SMAD3 co-binds with MYOD1 (Mullen et al., 2011). The overlap between MYOD1 and SMAD3 target genes demonstrate the significant association between both genes in skeletal muscle, in agreement with the tissue-specific TGFB1 signaling response (Mullen et al., 2011). The overlap percentage between our predicted binding sites and MYOD1 Chip-seq data (18 and 24.5%) confirms previous analyses in mice where they reported

only 20% of experimental validated distal enhancers in mouse myotubes with a bHLH (MyoD1) binding were actually bound by MYOD1 ChIP-seq data (Blum et al., 2012). Thus, suggesting that additional transcription-factors and/or histone modifications have a key role in MYOD1 binding. The SMAD3/MYOD1 co-bound regions for known target genes are also captured, such as the promoter regions of *ACTA1* and *ANKRD1*, both genes involved in skeletal muscle differentiation. We also demonstrated predicted MYOD1 binding regions are enriched for muscle regulatory regions across species (human, mouse and cow).

Altogether, we showed that co-expressed genes with *TGFB1* are enriched for SMAD3/MYOD1 binding sites, which we validated at the gene and enhancer level by proving not only MYOD1 and SMAD3 binding, but also their accessibility, in human, mouse and cow. In pigs, increased feed efficiency is associated with stimulation of muscle growth by TGFB1 signaling pathway (Jing et al., 2015). Finally, although not directly co-expressed with *TGFB1*, oxytocin (*OXT*) was DE in muscle and despite the lack of knowledge about its role in this tissue, previous work in cattle showed a massive increase of *OXT* expression in the muscle of bovines chronically exposed to anabolic steroids (De Jager et al., 2011). It is not known yet if oxytocin alone has an anabolic activity, but in a context where muscle growth seems to be associated with high FE animals, this hormone should be the focus of further investigation.

From the 13 regulator genes that are DE between groups, six are involved in respiratory chain and are up-regulated in HFE group. Genes *ND1, ND4, ND4L, ND5, ND6* and also *ND2* (which is DE but not identified as a regulator) are core subunits of the mitochondrial membrane respiratory chain Complex I (CI) which functions in the transfer of electrons from NADH to the respiratory chain, while ATP8 is part of Complex V and produces ATP from ADP in the presence of the proton gradient across the membrane. Interestingly, greater quantities of mitochondrial CI protein were associated with high FE cattle by Ramos and Kerley (2013) whereas Davis et al. (2016) found higher CI-CII and CI-CIII concentration ratios for the same group. Other studies demonstrated that HFE animals consume less oxygen (Chaves et al., 2015) and present lower plasma CO_2 concentrations, which suggests a decreased oxidation process (Gonano et al., 2014). In general, the literature suggests mitochondrial ADP has greater control of oxidative phosphorylation in high FE individuals (Lancaster et al., 2014) and their increased mitochondrial function may contribute to feed efficiency (Connor et al., 2010). In pigs, differences in mitochondrial function were reported when analyzing muscle (Vincent et al., 2015), blood (Liu et al., 2016) and adipose tissue transcriptomes (Louveau et al., 2016). Differences in metabolic rate associated with FE have long been discussed (Herd and Arthur, 2009) and here the hypothesis is corroborated by the up-regulation of *TSHB* in HFE animals, which stimulates production of T3 and T4 in thyroid, thus increasing metabolism. Metabolism is inhibited by *SST*, a down-regulated hormone in this group which was also found to be differentially connected between HFE and LFE.

Examining the DE genes, many hormones can be identified. Hormones are signaling proteins that are transported by the circulatory system to target distant organs in order to regulate physiology. Regarding the relationship between FE and other production traits of economic importance, FSHB, responsible for spermatozoa production by activating Sertoli cells in the testicles (Walker and Cheng, 2005), is up-regulated in LFE group and is inhibited by follistatin (FST), a gene found to be down-regulated in the same group. Moreover, in rats, it has been demonstrated that FSH secretion is stimulated by somatostatin expression, which is up-regulated in LFE animals (Kitaoka et al., 1989). In this scenario, one could argue that selection for high FE delay reproduction traits, something that could be related to the lower fat deposition in this group, as previously observed (Gomes et al., 2012; Santana et al., 2012; Alexandre et al., 2015). Indeed, differences in body composition and in intermediary metabolism can impact on reproductive traits (Shaffer et al., 2011) and it has been observed before that feed efficient bulls present features of delayed sexual maturity, i.e., decreased progressive motility of the sperm and higher abundance of tail abnormalities (Fontoura et al., 2016; Montanholi et al., 2016). Moreover, high FE heifers presented lower fat deposition and later sexual maturity which results in calving later in the calving season than their low FE counterparts (Shaffer et al., 2011; Randel and Welsh, 2013). LFE animals also exhibit down-regulation of *AMH* and the decrease of this hormone in serum is an excellent marker of Sertoli cells pubertal development (Rey et al., 1993).

Concerning the differences in lipid metabolism in divergent FE phenotypes, FGF21, a hormone up-regulated in liver of HFE animals, is associated in humans to decrease in body weight, blood triglycerides and LDL-cholesterol, with improvement in insulin sensitivity (Cheung and Deng, 2014). It is an hepatokine released to the bloodstream and an important regulator of lipid and glucose metabolism (Giralt et al., 2015). When we performed an enrichment analysis of co-expressed genes with *FGF21*, we indeed found terms related to plasma lipoprotein particle remodeling, regulation of lipoprotein oxidation and cholesterol efflux mostly due to *FGF21* co-expression with the apolipoproteins APOA4, APOC3 and APOM. In the same context, pro-melanin-concentrating hormone (PMCH) encodes three neuropeptides: neuropeptide-glycine-glutamic acid, neuropeptide-glutamic acid-isoleucine and melanin-concentrating hormone (MCH), the last one being the most extensively studied (Helgeson and Schmutz, 2008). MCH up-regulation has been related to obesity and insulin resistance, as well as increased appetite and reduced metabolism in murine models (Ludwig et al., 2001; Ito et al., 2003). The *PMCH* gene is up-regulated in LFE animals and harbors SNPs found to be associated with higher carcass fat levels and marbling score (Helgeson and Schmutz, 2008; Walter et al., 2014).

In this work, we were able to identify several biological processes known to be related to feed efficiency, which together with the validation of the main transcription factors of the network, demonstrate the quality of the data and the robustness of the analyses, giving us the confidence to identify candidate genes as regulators or biomarkers of superior animals for this trait. The regulatory genes *NR2F6* and *TGFB1* play central roles in liver and muscle, respectively, by regulating genes related to inflammatory response and muscle development and growth, two main biological mechanisms associated to feed efficiency.

Likewise, hormones and other proteins secreted in plasma as oxytocin, adrenomedulin, TSH, somatostatin, follistatin and AMH are interesting molecules to be explored as potential biomarkers of feed efficiency.

MATERIALS AND METHODS

Phenotypic Data and Biological Sample Collection

All animal protocols were approved by the Institutional Animal Care and Use Committee of Faculty of Food Engineering and Animal Sciences, University of São Paulo (FZEA-USP – protocol number 14.1.636.74.1). All procedures to collect phenotypes and biological samples were carried out at FZEA-USP, Pirassununga, State of São Paulo, Brazil. Ninety-eight Nellore bulls (16 to 20 months old and 376 ± 29 kg BW) were evaluated in a feeding trial comprised of 21 days of adaptation to feedlot diet and place and a 70-day period of data collection. Total mixed ration was offered *ad libitum* and daily dry matter intake (DMI) was individually measured. Animals were weighed at the beginning, at the end and every 2 weeks during the experimental period. Feed efficiency was estimated by RFI which is the residual of the linear regression that estimates DMI based on average daily gain and mid-test metabolic body weight (Koch et al., 1963). 40 animals selected either as high feed efficiency (HFE) or low feed efficiency (LFE) groups were slaughtered on 2 days with a 6-day interval. Adrenal gland (longitudinal section), hypothalamus, liver (lateral portion of the left lobe), skeletal muscle (medial portion of *Longissimus lumborum,* close to 12th rib) and pituitary samples were collected from each animal, rapidly frozen in liquid nitrogen and stored at –80°C. Further information about management and phenotypic measures of the animals used in this study can be found in Alexandre et al. (2015).

RNAseq Data Generation

Samples of nine animals from each feed efficiency group (high and low) were selected for RNAseq using RFI measure. For hypothalamus and pituitary, the nitrogen frozen tissue was macerated with crucibles and pistils to ensure all portions of the tissue were represented, and stored in aliquots at –80°C. Then, RNA was extracted using AllPrep DNA/RNA/Protein Mini kit (QIAGEN, Crawley, United Kingdom). For liver, muscle and adrenal gland, a cut was made in the frozen tissue and the RNA was extracted using RNeasy Mini Kit (QIAGEN, Crawley, United Kingdom). RNA quality and quantity were assessed using automated capillary gel electrophoresis on a Bioanalyzer 2100 with RNA 6000 Nano Labchips according to the manufacturer's instructions (Agilent Technologies Ireland, Dublin, Ireland). Samples that presented an RNA integrity number (RIN) of less than 8.0 were discarded.

RNA libraries were constructed using the TruSeq^{TM} Stranded mRNA LT Sample Prep Protocol and sequenced on Illumina HiSeq 2500 equipment in a HiSeq Flow Cell v4 using the HiSeq SBS Kit V4 (2×100 pb). Liver, pituitary and hypothalamus were sequenced on the same run, each one in a different lane. Muscle and adrenal gland were sequenced in a second run, in different lanes.

Gene Expression Estimation

The quality of the sequencing was evaluated using the software FastQC Version 3[1]. Sequence alignment against the bovine reference genome (UMD3.1) was performed using STAR Version 2.2.1 (Dobin et al., 2013), according to the standard parameters and including the annotation file (Ensembl release 89) and secondary alignments, duplicated reads and reads failing vendor quality checks were removed using Samtools Version 1.9 (Li et al., 2009). Then, HTseq Version 0.6.0 (Anders et al., 2014) was used to generate gene read counts and expression values were estimated by fragments per kilobase of gene per million mapped reads (FPKM). Genes with average value lower than 0.2 FPKM across all samples and tissues were discarded.

Gene expression normalization was performed using the following mixed effect model (Reverter et al., 2005):

$$Y_{ijkl} = \mu + L_i + G_j + GT_{jk} + GP_{jl} + e_{ijkl}$$

where, the log2-transformed FPKM value for i-th library (86 levels), j-th gene (17,354 levels), k-th tissue (5 levels), l-th RFI phenotype (2 levels), corresponding to Y_{ijkl}, was modeled as a function of the fixed effect of library (L_i) and the random effects of gene (G_j), gene by tissue (GT_{jk}) and gene by RFI phenotype (GP_{jl}). Random residual (e_{ijkl}) was assumed to be independent and identically distributed. Variance component estimates and solutions to the model were obtained using VCE6 (Groeneveld et al., 2010). Normalized mean expression (NME) values for each gene were defined as the linear combination of the solutions for random effects.

The mixed model used to normalize the expression data explained 96% of the variation in gene expression, of which the largest proportion (0.30) was due to tissue-specificity. Contrariwise, differences between HFE and LFE represented no variation (0.27×10^{-11}). For that reason, normalized mean expression (NME) was only used to identify tissue specific genes and the raw FPKM values were used for differential expression and co-expression analysis.

Gene Selection for Network Construction

In order to select a set of relevant genes for network analysis, we defined five categories based on the following inclusion criteria:

Differential Expression (DE)

The mean expression value of each gene, for each group (HFE and LFE) and each tissue was calculated and then the expression of LFE group was subtracted from the expression in HFE group. Next, genes were ranked according to their mean expression in all samples for each tissue and divided into five bins. Genes were considered differentially expressed when the difference between the expression in HFE and LFE groups were greater than 3.1 or smaller than – 3.1 standard deviation from the mean in each bin, corresponding to a t-test $P < 0.001$ (Weber et al., 2016).

[1]http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Harboring SNPs

Genes harboring SNPs associated with feed efficiency, mainly indicated by GWAS, were identified using the PubMed database[2] and the AnimalQTL database – Release 35[3] and only bovine data were considered regardless of breed.

Tissue Specific (TS)

A gene was considered as tissue specific when the average NME in that tissue was greater than one standard deviation from the mean of all genes and the average NME in all the other four tissues was smaller than zero.

Secreted

The human secretome database[4] (Uhlén et al., 2005; Uhlen et al., 2015) was used to select genes encoding proteins secreted in plasma by any of the analyzed tissues (adrenal gland, hypothalamus, liver, muscle and pituitary).

Key Regulators

In order to identify key regulatory genes to be included in the co-expression network, a list of genes were obtained from the Animal Transcription Factor Database 2.0[5] (Zhang et al., 2015) and it was compared to a set of potential target genes in each tissue, composed of the categories: TS, DE, harboring SNPs and secreted. The analysis was based on regulatory impact factor metrics (Reverter et al., 2010), which comprises a set of two metrics designed to assign scores to regulator genes consistently differentially co-expressed with target genes and to those with the most altered ability to predict the abundance of target genes. Those scores deviating $\pm$ 1.96 standard deviation from the mean (corresponding to $P < 0.05$) were considered significant. Genes presenting mean expression value less than the mean of all genes expressed were not considered in this analysis.

Some of the genes selected by the categories above were represented by more than one Ensembl ID. Those duplications were removed for further analysis, keeping only the expression value of the most meaningful Ensembl ID. Additionally, genes with mean expression across the samples equal to zero were also removed from further analysis.

Co-expression Network Analysis

For gene network inference, genes selected using the five categories described previously were used as nodes and significant connections (edges) between them were identified using the Partial Correlation and Information Theory (PCIT) algorithm (Reverter and Chan, 2008), considering all animals and all tissues. PCIT determinates the significance of the correlation between two nodes after accounting for all the other nodes in the network. Connections between gene nodes were accepted when the partial correlation was greater than two standard deviations from the mean ($P < 0.01$). The output of PCIT was visualized on Cytoscape Version 3.6.1 (Shannon et al., 2003).

[2]www.ncbi.nlm.nih.gov/pubmed/

[3]www.animalgenome.org/cgi-bin/QTLdb/index

[4]www.proteinatlas.org/humanproteome/secretome

[5]http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/

Network Validation Through Transcription Factor Biding Motifs Analysis

Using the regulatory impact factor metric (RIF) we prioritized key regulator genes from gene expression data and predicted target genes based on co-expression network. In order to assess whether those target genes were enriched for motifs associated with the top most connected regulators in the network with a DNA binding domain (transcription factors – TF), we performed motif discovery analysis in the set of co-expressed target genes (first neighbors of the TF) using the i-cistarget method (Herrmann et al., 2012) and i-Regulon v1.3, a Cytoscape plug-in (Janky et al., 2014). These tools use humans (hg19) as the reference species, therefore only genes with human orthology are assessed. Then, to validate the binding of the identified genomic regions by the TFs, we performed a region enrichment analysis across experimentally available TF bound regions from ChiP-seq in cell lines from the ENCODE consortium (1,394 TF binding site tracks, Encode Project Consortium, 2012; Sloan et al., 2016). Briefly, the tools evaluate whether there is an over-representation of motifs in the set of co-expressed genes and across evolution. We examined 10 kb upstream of the gene transcription start site and their conservation in 7 vertebrate species, including cow. Thus, the tools provide over-represented motifs across evolution, allowing us to predict regulatory interactions TF to target gene in cow. In our analysis, we performed motif discovery using i-Regulon v1.3 (Janky et al., 2014) and i-cistarget database 3 (Herrmann et al., 2012), that is using their 9713 motif collection. Both methods result in highly similar enrichments. Whereas i-Regulon is a user-friendly method to deliver a regulatory network, i-cistarget also yields the genomic position of the TF binding in the genome. Both i-Regulon and i-cistarget can be used to validate the TF binding on predicted genomic regions in the human genome. The tool contains a collection of TF ChIP-seq data in cell lines mostly from the ENCODE consortium (1,394 TF binding tracks), 2003 Histone modifications from the ENCODE consortium and Epigenomics roadmap and 908 Histone modification and open-chromatin. The tool allows to perform an enrichment of the different human tracks at the region level.

Finally, we converted identified enhancer regions into cow coordinates and searched for regions of open-chromatin using data from publicly available studies in cow tissues. Namely, cow liver promoters and enhancer from Villar et al. (2015) (Array Express Accession number E-MTAB-2633) and skeletal muscle cow promoters from Zhao et al. (2015) (GSE61936). For *MYOD1* and *SMAD3* binding in myotubes and pro-B cells, data from Mullen et al. (2011) was used (GEO: GSE21621); and for *MYOD1* binding in primary myotubes, data from Cao et al. (2010) (GEO: GSE20059) was used.

Differential Connectivity

In order to explore differentially connected genes between HFE and LFE, two networks were created, one for each condition, using the same methodology described before. Then, the number of connections of each gene in each condition was computed and

scaled so that connectivity varied from 0 to 1, making it possible to compare the same gene in the two networks. The connectivity in LFE group was subtracted from the connectivity in HFE group and results deviating ± 1.96 standard deviation from the mean were considered significant ($P < 0.05$).

Functional Enrichment

Functional enrichment analysis was performed on the online platform GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool[6]), using all genes that passed FPKM filter as background, hypergeometric test and multiple test correction (FDR – false discovery rate). The human database was used to take advantage of a more comprehensive knowledgebase regarding gene functions. GO terms were considered significant when Padj < 0.05. For genes in co-expression networks, visualized using Cytoscape (Shannon et al., 2003), the functional enrichment was performed with BiNGO plug-in (Maere et al., 2005) using the same background genes and statistical test.

DATA AVAILABILITY

Datasets supporting the results of this article are public available in the European Nucleotide Archive (ENA) as part of FAANG consortium under de study ID PRJEB27337 and can be accessed following the link https://www.ebi.ac.uk/ena/data/view/PRJEB27337. Moreover, any additional information as well as the scripts used to perform the analysis are available upon request, please email pamela.alexandre@usp.br. This manuscript has been published as a preprint in bioRxiv, doi: https://doi.org/10.1101/360396.

[6] http://cbl-gorilla.cs.technion.ac.il/

AUTHOR CONTRIBUTIONS

PA performed formal analysis, investigation and visualization of all presented data and wrote the original draft. HF was the overall project leader responsible for conceptualization, project administration and supervision of PA. AR performed bioinformatics analysis and investigation and supervised PA. MN-S performed bioinformatics analysis and investigation. JF was responsible for funding acquisition and project administration. LP-N provided computing resources and conceptualization. All authors contributed to reviewing and editing of this manuscript's final version.

FUNDING

This study and PA scholarships were funded by São Paulo Research Foundation (FAPESP) – Proc. 2014/07566-2, 2014/02493-7, 2015/22276-3 and 2017/14707-0. MN-S was funded by the CSIRO Science Excellence Research Office.

ACKNOWLEDGMENTS

The authors thank Mrs. Elisângela C. M. Oliveira and Ms. Arina L. Rochetti for all technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00230/full#supplementary-material

REFERENCES

Aerts, S., Quan, X. J., Claeys, A., Sanchez, M. N., Tate, P., Yan, J., et al. (2010). Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in drosophila uncovers a regulatory basis for sensory specification. *PLoS Biol.* 8:e1000435. doi: 10.1371/journal.pbio.1000435.g001

Alexandre, P. A., Kogelman, L. J. A., Santana, M. H. A., Passarelli, D., Pulz, L. H., Fantinato-Neto, P., et al. (2015). Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics* 16:1073. doi: 10.1186/s12864-015-2292-8

Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Archer, J. A., Arthur, P. F., Herd, R. M., Parnell, P. F., and Pitchford, W. S. (1997). Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *J. Anim. Sci.* 75, 2024–2032. doi: 10.2527/1997.7582024x

Archer, J. A., Richardson, E. C., Herd, R. M., and Arthur, P. F. (1999). Potential for selection to improve efficiency of feed use in beef cattle: a review. *Aust. J. Agric. Res.* 50, 147–161. doi: 10.1071/A98075

Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B. D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Genes Dev.* 26, 2763–2779. doi: 10.1101/gad.200113.112

Bolormaa, S., Hayes, B. J., Savin, K., Hawken, R., Barendse, W., Arthur, P. F., et al. (2011). Genome-wide association studies for feedlot and growth traits in cattle. *J. Anim. Sci.* 89, 1684–1697. doi: 10.2527/jas.2010-3079

Cafe, L. M., Robinson, D. L., Ferguson, D. M., Geesink, G. H., and Greenwood, P. L. (2011). Temperament and hypothalamic-pituitary-adrenal axis function are related and combine to affect growth, efficiency, carcass, and meat quality traits in Brahman steers. *Domest. Anim. Endocrinol.* 40, 230–240. doi: 10.1016/j.domaniend.2011.01.005

Cao, P., Hanai, J.-I., Tanksale, P., Imamura, S., Sukhatme, V. P., and Lecker, S. H. (2009). Statin-induced muscle damage and atrogin-1 induction is the result of a geranylgeranylation defect. *FASEB J.* 23, 2844–2854. doi: 10.1096/fj.08-128843

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G. J., Parker, M. H., et al. (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell.* 18, 662–674. doi: 10.1016/j.devcel.2010.02.014

Chaves, A. S., Nascimento, M. L., Tullio, R. R., Rosa, A. N., Alencar, M. M., and Lanna, D. P. (2015). Relationship of efficiency indices with performance, heart rate, oxygen consumption, blood parameters, and estimated heat production in Nellore steers. *J. Anim. Sci.* 93, 5036–5046. doi: 10.2527/jas.2015-9066

Chen, L., Mao, F., Crews, D. H., Vinsky, M., and Li, C. (2014). Phenotypic and genetic relationships of feeding behavior with feed intake, growth performance, feed efficiency, and carcass merit traits in Angus and Charolais steers. *J. Anim. Sci.* 92, 974–983. doi: 10.2527/jas.2013-6926

Cheung, B. M., and Deng, H. B. (2014). Fibroblast growth factor 21: a promising therapeutic target in obesity-related diseases. *Exp. Rev. Cardiovasc. Ther.* 12, 659–666. doi: 10.1586/14779072.2014.904745

Connor, E. E., Kahl, S., Elsasser, T. H., Parker, J. S., Li, R. W., Van Tassell, C. P., et al. (2010). Enhanced mitochondrial complex gene function and reduced liver size may mediate improved feed efficiency of beef cattle during compensatory growth. *Funct. Integr. Genomics* 10, 39–51. doi: 10.1007/s10142-009-0138-7

Davis, M. P., Brooks, M. A., and Kerley, M. S. (2016). Relationship between residual feed intake and lymphocyte mitochondrial complex protein concentration and ratio in crossbred steers. *J. Anim. Sci.* 94, 1587–1591. doi: 10.2527/jas2015-9843

De Jager, N., Hudson, N. J., Reverter, A., Wang, Y., Nagaraj, S. H., Cafe, L. M., et al. (2011). Chronic exposure to anabolic steroids induces the muscle expression of oxytocin and a more than fiftyfold increase in circulating oxytocin in cattle. *Physiol. Genomics* 43, 467–478. doi: 10.1152/physiolgenomics.00226.2010

de la Fuente, A. (2010). From "differential expression" to "differential networking" - identification of dysfunctional regulatory networks in diseases. *Trends Genet.* 26, 326–333. doi: 10.1016/j.tig.2010.05.001

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Encode Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Fontoura, A. B. P., Montanholi, Y. R., Diel de Amorim, M., Foster, R. A., Chenier, T., and Miller, S. P. (2016). Associations between feed efficiency, sexual maturity and fertility-related measures in young beef bulls. *Animal* 10, 96–105. doi: 10.1017/S1751731115001925

Foote, A. P., Tait, R. G., Keisler, D. H., Hales, K. E., and Freetly, H. C. (2016). Leptin concentrations in finishing beef steers and heifers and their association with dry matter intake, average daily gain, feed efficiency, and body composition. *Domest. Anim. Endocrinol.* 55, 136–141. doi: 10.1016/j.domaniend.2015.12.007

Francisco, C. L., Resende, F. D., Benatti, J. M. B., Castilhos, A. M., Cooke, R. F., and Jorge, A. M. (2015). Impacts of temperament on Nellore cattle: physiological responses, feedlot performance, and carcass characteristics. *J. Anim. Sci.* 93:5419. doi: 10.2527/jas.2015-9411

Gerber, P. J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., et al. (2013). *Tackling Climate Change Through Livestock - A Global Assessment of Emissions and Mitigation Opportunities*. Rome: Food & Agriculture Organization.

Giralt, M., Gavaldà-Navarro, A., and Villarroya, F. (2015). Fibroblast growth factor-21, energy balance and obesity. *Mol. Cell. Endocrinol.* 418, 66–73. doi: 10.1016/j.mce.2015.09.018

Gomes, R. C., Sainz, R. D., Silva, S. L., César, M. C., Bonin, M. N., and Leme, P. R. (2012). Feedlot performance, feed efficiency reranking, carcass traits, body composition, energy requirements, meat quality and calpain system activity in Nellore steers with low and high residual feed intake. *Livest. Sci.* 150, 265–273. doi: 10.1016/j.livsci.2012.09.012

Gonano, C., Montanholi, Y., Schenkel, F., Smith, B., Cant, J., and Miller, S. (2014). The relationship between feed efficiency and the circadian profile of blood plasma analytes measured in beef heifers at different physiological stages. *Animal* 8, 1684–1698. doi: 10.1017/S1751731114001463

Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., Causeur, D., et al. (2017). A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. *BMC Genomics* 18:244. doi: 10.1186/s12864-017-3639-0

Groeneveld, E., Kovač, M., and Mielenz, N. (2010). *VCE User's Guide and Reference Manual, Version 6.0*. Neustadt: Institute of Farm Animal Genetics.

Hayes, B. J., Lewin, H. A., and Goddard, M. E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29, 206–214. doi: 10.1016/j.tig.2012.11.009

Helgeson, S. C., and Schmutz, S. M. (2008). Genetic variation in the pro-melanin-concentrating hormone gene affects carcass traits in *Bos taurus* cattle. *Anim. Genet.* 39, 310–315. doi: 10.1111/j.1365-2052.2008.01717.x

Herd, R. M., and Arthur, P. F. (2009). Physiological basis for residual feed intake. *J. Anim. Sci.* 87, E64–E71. doi: 10.2527/jas.2008-1345

Herd, R. M., Oddy, V. H., and Richardson, E. C. (2004). Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. *Aust. J. Exp. Agric.* 44, 423–430. doi: 10.1071/ea02220

Hermann-Kleiter, N., and Baier, G. (2014). Orphan nuclear receptor NR2F6 acts as an essential gatekeeper of Th17 CD4+ T cell effector functions. *Cell Commun. Signal.* 12, 1–12. doi: 10.1186/1478-811X-12-38

Hermann-Kleiter, N., Gruber, T., Lutz-Nicoladoni, C., Thuille, N., Fresser, F., Labi, V., et al. (2008). The nuclear orphan receptor NR2F6 suppresses lymphocyte activation and T Helper 17-dependent autoimmunity. *Immunity* 29, 205–216. doi: 10.1016/j.immuni.2008.06.008

Herrmann, C., Van De Sande, B., Potier, D., and Aerts, S. (2012). i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* 40:e114. doi: 10.1093/nar/gks543

Ito, M., Gomori, A., Ishihara, A., Oda, Z., Mashiko, S., Matsushita, H., et al. (2003). Characterization of MCH-mediated obesity in mice. *Am. J. Physiol. Endocrinol. Metab.* 284, E940–E945. doi: 10.1152/ajpendo.00529.2002

Janky, R., Verfaillie, A., Imrichová, H., van de Sande, B., Standaert, L., Christiaens, V., et al. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* 10:e1003731. doi: 10.1371/journal.pcbi.1003731

Jing, L., Hou, Y., Wu, H., Miao, Y., Li, X., Cao, J., et al. (2015). Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential residual feed intake in pigs. *Sci. Rep.* 5:11953. doi: 10.1038/srep11953

Karisa, B., Moore, S., and Plastow, G. (2014). Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. *Anim. Sci. J.* 85, 374–387. doi: 10.1111/asj.12159

Kelly, A. K., McGee, M., Crews, D. H., Fahey, A. G., Wylie, A. R., and Kenny, D. A. (2010). Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. *J. Anim. Sci.* 88, 109–123. doi: 10.2527/jas.2009-2196

Kitaoka, M., Takano, K., Kojima, I., and Ogata, E. (1989). A stimulatory effect of somatostatin: enhancement of activin A-mediated FSH secretion in rat pituitary cells. *Biochem. Biophys. Res. Commun.* 162, 958–962. doi: 10.1016/0006-291X(89)90766-3

Klepsch, V., Hermann-Kleiter, N., and Baier, G. (2016). Beyond CTLA-4 and PD-1: orphan nuclear receptor NR2F6 as T cell signaling switch and emerging target in cancer immunotherapy. *Immunol. Lett.* 178, 31–36. doi: 10.1016/j.imlet.2016.03.007

Koch, R. M., Swiger, L. A., Chambers, D., and Gregory, K. E. (1963). Efficiency of feed use in beef cattle. *J. Anim. Sci.* 22, 486–494. doi: 10.2527/jas1963.222486x

Lancaster, P. A., Carstens, G. E., Michal, J. J., Brennan, K. M., Johnson, K. A., and Davis, M. E. (2014). Relationships between residual feed intake and hepatic mitochondrial function in growing beef cattle. *J. Anim. Sci.* 92, 3134–3141. doi: 10.2527/jas.2013-7409

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, D., Black, B. L., and Derynck, R. (2001). TGF-beta inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. *Genes Dev.* 15, 2950–2966. doi: 10.1101/gad.925901

Liu, H., Nguyen, Y. T., Nettleton, D., Dekkers, J. C. M., and Tuggle, C. K. (2016). Post-weaning blood transcriptomic differences between Yorkshire pigs divergently selected for residual feed intake. *BMC Genomics* 17:73. doi: 10.1186/s12864-016-2395-x

Louveau, I., Vincent, A., Tacher, S., Gilbert, H., and Gondret, F. (2016). Increased expressions of genes and proteins involved in mitochondrial oxidation and antioxidant pathway in adipose tissue of pigs selected for a low residual feed intake. *J. Anim. Sci.* 94, 5042–5054. doi: 10.2527/jas2016-0619

Ludwig, D. S., Tritos, N. A., Mastaitis, J. W., Kulkarni, R., Kokkotou, E., Elmquist, J., et al. (2001). Melanin-concentrating hormone overexpression in transgenic mice leads to obesity and insulin resistance. *J. Clin. Invest.* 107, 379–386. doi: 10.1172/JCI10660

Mader, C. J., Montanholi, Y. R., Wang, Y. J., Miller, S. P., Mandell, I. B., McBride, B. W., et al. (2009). Relationships among measures of growth performance and efficiency with carcass traits, visceral organ mass, and pancreatic digestive enzymes in feedlot cattle. *J. Anim. Sci.* 87, 1548–1557. doi: 10.2527/jas.2008-0914

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Mani, V., Harris, A. J., Keating, A. F., Weber, T. E., Dekkers, J. C. M., and Gabler, N. K. (2013). Intestinal integrity, endotoxin transport and detoxification in

pigs divergently selected for residual feed intake. *J. Anim. Sci.* 91, 2141–2150. doi: 10.2527/jas.2012-6053

Montanholi, Y. R., Fontoura, A. B. P., Diel de Amorim, M., Foster, R. A., Chenier, T., and Miller, S. P. (2016). Seminal plasma protein concentrations vary with feed efficiency and fertility-related measures in young beef bulls. *Reprod. Biol.* 16, 147–156. doi: 10.1016/j.repbio.2016.04.002

Montanholi, Y. R., Swanson, K. C., Palme, R., Schenkel, F. S., McBride, B. W., Lu, D., et al. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* 4, 692–701. doi: 10.1017/S1751731109991522

Montanholi, Y. R., Swanson, K. C., Schenkel, F. S., McBride, B. W., Caldwell, T. R., and Miller, S. P. (2009). On the determination of residual feed intake and associations of infrared thermography with efficiency and ultrasound traits in beef bulls. *Livest. Sci.* 125, 22–30. doi: 10.1016/j.livsci.2009.02.022

Mota, L. F. M., Bonafé, C. M., Alexandre, P. A., Santana, M. H., Novais, F. J., Toriyama, E., et al. (2017). Circulating leptin and its muscle gene expression in Nellore cattle with divergent feed efficiency. *J. Anim. Sci. Biotechnol.* 8:71. doi: 10.1186/s40104-017-0203-3

Mukiibi, R., Vinsky, M., Keogh, K. A., Fitzsimmons, C., Stothard, P., Waters, S. M., et al. (2018). Transcriptome analyses reveal reduced hepatic lipid synthesis and accumulation in more feed efficient beef cattle. *Sci. Rep.* 8:7303. doi: 10.1038/s41598-018-25605-3

Mullen, A. C., Orlando, D. A., Newman, J. J., Lovén, J., Kumar, R. M., Bilodeau, S., et al. (2011). Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* 147, 565–576. doi: 10.1016/j.cell.2011.08.050

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A., and Freetly, H. C. (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10:e0129174. doi: 10.1371/journal.pone.0129174

Myer, P. R., Wells, J. E., Smith, T. P. L., Kuehn, L. A., and Freetly, H. C. (2016). Microbial community profiles of the jejunum from steers differing in feed efficiency. *J. Anim. Sci.* 94:327. doi: 10.2527/jas.2015-9839

Naval-Sańchez, M., Potier, D., Haagen, L., Sánchez, M., Munck, S., Van De Sande, B., et al. (2013). Comparative motif discovery combined with comparative transcriptomics yields accurate targetome and enhancer predictions. *Genome Res.* 23, 74–88. doi: 10.1101/gr.140426.112

Nkrumah, J. D., Keisler, D. H., Crews, D. H., Basarab, J. A., Wang, Z., Li, C., et al. (2007). Genetic and phenotypic relationships of serum leptin concentration with performance, efficiency of gain, and carcass merit of feedlot cattle. *J. Anim. Sci.* 85, 2147–2155. doi: 10.2527/jas.2006-764

Novais, F. J., Pires, P. R. L., Alexandre, P. A., Dromms, R. A., Iglesias, A. H., Ferraz, J. B. S., et al. (2019). Identification of a metabolomic signature associated with feed efficiency in beef cattle. *BMC Genomics* 20:8. doi: 10.1186/s12864-018-5406-2

Oliveira, P. S. N., Cesar, A. S. M., Nascimento, M. L., Chaves, A. S., Tizioto, P. C., Tullio, R. R., et al. (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet.* 15:100. doi: 10.1186/s12863-014-0100-0

Paradis, F., Yue, S., Grant, J. R., Stothard, P., Basarab, J. A., and Fitzsimmons, C. (2015). Transcriptomic analysis by RNA sequencing reveals that hepatic interferon-induced genes may be associated with feed efficiency in beef heifers. *J. Anim. Sci.* 93:3331. doi: 10.2527/jas.2015-8975

Pereira, F. A., Tsai, M. J., and Tsai, S. Y. (2000). COUP-TF orphan nuclear receptors in development and differentiation. *Cell Mol. Life Sci.* 57, 1388–1398. doi: 10.1007/PL00000624

Potier, D., Davie, K., Hulselmans, G., NavalSanchez, M., Haagen, L., Huynh-Thu, V. A., et al. (2014). Mapping gene regulatory networks in drosophila eye development by large-scale transcriptome perturbations and motif inference. *Cell Rep.* 9, 2290–2303. doi: 10.1016/j.celrep.2014.11.038

Ramayo-Caldas, Y., Ballester, M., Sánchez, J. P., González-Rodríguez, O., Revilla, M., Reyer, H., et al. (2018). Integrative approach using liver and duodenum RNA-Seq data identifies candidate genes and pathways associated with feed efficiency in pigs. *Sci. Rep.* 8:558. doi: 10.1038/s41598-017-19072-5

Ramos, M. H., and Kerley, M. S. (2013). Mitochondrial complex I protein differs among residual feed intake phenotype in beef cattle. *J. Anim. Sci.* 91, 3299–3304. doi: 10.2527/jas.2012-5589

Randel, R. D., and Welsh, T. H. (2013). Joint Alpharma-Beef species symposium: interactions of feed efficiency with beef heifer reproductive development. *J. Anim. Sci.* 91, 1323–1328. doi: 10.2527/jas.2012-5679

Reverter, A., Barris, W., McWilliam, S., Byrne, K. A., Wang, Y. H., Tan, S. H., et al. (2005). Validation of alternative methods of data normalization in gene co-expression studies. *Bioinformatics* 21, 1112–1120. doi: 10.1093/bioinformatics/bti124

Reverter, A., and Chan, E. K. F. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. doi: 10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Pérez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26, 896–904. doi: 10.1093/bioinformatics/btq051

Rey, R., Lordereau-Richard, I., Carel, J. C., Barbet, P., Cate, R. L., Roger, M., et al. (1993). Anti-müllerian hormone and testosterone serum levels are inversely during normal and precocious pubertal development. *J. Clin. Endocrinol. Metab.* 77, 1220–1226. doi: 10.1210/jcem.77.5.8077315

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–329. doi: 10.1038/nature14248

Rolf, M. M., Taylor, J. F., Schnabel, R. D., McKay, S. D., McClure, M. C., Northcutt, S. L., et al. (2011). Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* 43, 367–374. doi: 10.1111/j.1365-2052.2011.02273.x

Saatchi, M., Beever, J. E., Decker, J. E., Faulkner, D. B., Freetly, H. C., Hansen, S. L., et al. (2014). QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. *BMC Genomics* 15:1004. doi: 10.1186/1471-2164-15-1004

Santana, M. H. A., Rossi, P., Almeida, R., and Cucco, D. C. (2012). Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. *Livest. Sci.* 145, 252–257. doi: 10.1016/j.livsci.2012.02.012

Santana, M. H. A., Utsunomiya, Y. T., Neves, H. H. R., Gomes, R. C., Garcia, J. F., Fukumasu, H., et al. (2014). Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *BMC Genet.* 15:21. doi: 10.1186/1471-2156-15-21

Schmierer, B., and Hill, C. S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* 8, 970–982. doi: 10.1038/nrm2297

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18:386. doi: 10.1186/s12864-017-3754-y

Shaffer, K. S., Turk, P., Wagner, W. R., and Felton, E. E. D. (2011). Residual feed intake, body composition, and fertility in yearling beef heifers. *J. Anim. Sci.* 89, 1028–1034. doi: 10.2527/jas.2010-3322

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software Environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shin, J. H., Li, R. W., Gao, Y., Baldwin, R. VI, and Li, C. J. (2012). Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. *Funct. Integr. Genomics* 12, 119–130. doi: 10.1007/s10142-012-0263-6

Shindo, T., Kurihara, H., Kurihara, Y., Morita, H., and Yazaki, Y. (1998). Upregulation of endothelin-1 and adrenomedullin gene expression in the mouse endotoxin shock model. *J. Cardiovasc. Pharmacol.* 31(Suppl. 1), S541–S544. doi: 10.1097/00005344-199800001-00156

Shoji, H., Minamino, N., Kangawa, K., and Matsuo, H. (1995). Endotoxin markedly elevates plasma concentration and gene transcription of adrenomedullin in rat. *Biochem. Biophys. Res. Commun.* 215, 531–537. doi: 10.1006/bbrc.1995.2497

Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., et al. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Res.* 44, D726–D732. doi: 10.1093/nar/gkv1160

Swami, M. (2009). Networking complex traits. *Nat. Rev. Genet.* 10:2566. doi: 10.1038/ng.332

Tronche, F., and Yaniv, M. (1992). HNF1, a homeoprotein member of the hepatic transcription regulatory network. *BioEssays* 14, 579–587. doi: 10.1002/bies.950140902

Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A.-K., Amini, B., Andersen, E., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* 4, 1920–1932. doi: 10.1074/mcp.M500279-MCP200

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Tissue-based map of the human proteome. *Science* 347, 1260419–1260419. doi: 10.1126/science.1260419

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., et al. (2015). Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566. doi: 10.1016/j.cell.2015.01.006

Vincent, A., Louveau, I., Gondret, F., Tréfeu, C., Gilbert, H., and Lefaucheur, L. (2015). Divergent selection for residual feed intake affects the transcriptomic and proteomic profiles of pig skeletal muscle. *J. Anim. Sci.* 93, 2745–2758. doi: 10.2527/jas.2015-8928

Walker, W. H., and Cheng, J. (2005). FSH and testosterone signaling in Sertoli cells. *Reproduction* 130, 15–28. doi: 10.1530/rep.1.00358

Walter, L. J., Gasch, C. A., Mcevers, T. J., Hutcheson, J. P., Defoor, P., Marquess, F. L. S., et al. (2014). Association of pro-melanin concentrating hormone genotype with beef carcass quality and yield. *J. Anim. Sci.* 92, 325–331. doi: 10.2527/jas2013-6931

Weber, K. L., Welly, B. T., Van Eenennaam, A. L., Young, A. E., Port-Neto, L. R., Reverter, A., et al. (2016). Identification of Gene networks for residual feed intake in Angus cattle using genomic prediction and RNA-seq. *PLoS One* 11:e0152274. doi: 10.1371/journal.pone.0152274

Widmann, P., Reverter, A., Weikard, R., Suhre, K., Hammon, H. M., Albrecht, E., et al. (2015). Systems biology analysis merging phenotype, metabolomic and genomic data identifies Non-SMC condensin I complex, subunit G (NCAPG) and cellular maintenance processes as major contributors to genetic variability in bovine feed efficiency. *PLoS One* 10:e0124574. doi: 10.1371/journal.pone.0124574

Zarek, C. M., Lindholm-Perry, A. K., Kuehn, L. A., and Freetly, H. C. (2017). Differential expression of genes related to gain and intake in the liver of beef cattle. *BMC Res. Notes* 10:1. doi: 10.1186/s13104-016-2345-3

Zhang, H. M., Liu, T., Liu, C. J., Song, S., Zhang, X., Liu, W., et al. (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* 43, D76–D81. doi: 10.1093/nar/gku887

Zhao, C., Carrillo, J. A., Tian, F., Zan, L., Updike, S. M., Zhao, K., et al. (2015). Genome-wide H3K4me3 analysis in angus cattle with divergent tenderness. *PLoS One* 10:e0115358. doi: 10.1371/journal.pone.0115358

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Alexandre, Naval-Sanchez, Porto-Neto, Ferraz, Reverter and Fukumasu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 April 2019
doi: 10.3389/fgene.2019.00348

Genetic Regulation of Liver Metabolites and Transcripts Linking to Biochemical-Clinical Parameters

*Siriluck Ponsuksili[1], Nares Trakooljul[1], Frieder Hadlich[1], Karen Methling[2], Michael Lalk[2], Eduard Murani[1] and Klaus Wimmers[1,3]**

[1] Leibniz Institute for Farm Animal Biology (FBN), Institute for Genome Biology, Functional Genome Analysis Research Unit, Dummerstorf, Germany, [2] Institute for Biochemistry – Metabolomics, University of Greifswald, Greifswald, Germany, [3] Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany

Given the central metabolic role of the liver, hepatic metabolites and transcripts reflect the organismal physiological state. Biochemical-clinical plasma biomarkers, hepatic metabolites, transcripts, and single nucleotide polymorphism (SNP) genotypes of some 300 pigs were integrated by weighted correlation networks and genome-wide association analyses. Network-based approaches of transcriptomic and metabolomics data revealed linked of transcripts and metabolites of the pentose phosphate pathway (PPP). This finding was evidenced by using a NADP/NADPH assay and *HDAC4* and *G6PD* transcript quantification with the latter coding for first limiting enzyme of this pathway and by RNAi knockdown experiments of *HDAC4*. Other transcripts including *ARG2* and *SLC22A7* showed link to amino acids and biomarkers. The amino acid metabolites were linked with transcripts of immune or acute phase response signaling, whereas the carbohydrate metabolites were highly enrich in cholesterol biosynthesis transcripts. Genome-wide association analyses revealed 180 metabolic quantitative trait loci (mQTL) ($p < 10^{-4}$). Trans-4-hydroxy-L-proline ($p = 6 \times 10^{-9}$), being strongly correlated with plasma creatinine (CREA), showed strongest association with SNPs on chromosome 6 that had pleiotropic effects on *PRODH2* expression as revealed by multivariate analysis. Consideration of shared marker association with biomarkers, metabolites, and transcripts revealed 144 SNPs associated with 44 metabolites and 69 transcripts that are correlated with each other, representing 176 mQTL and expression quantitative trait loci (eQTL). This is the first work to report genetic variants associated with liver metabolite and transcript levels as well as blood biochemical-clinical parameters in a healthy porcine model. The identified associations provide links between variation at the genome, transcriptome, and metabolome level molecules with clinically relevant phenotypes. This approach has the potential to detect novel biomarkers displaying individual variation and promoting predictive biology in medicine and animal breeding.

Keywords: mQTL, eQTL, metabolite, transcript, SNPs, biochemical-clinical traits, biomarker, pig model

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Kirk L. Pappan, Metabolon, United States
Luiz Brito, Purdue University, United States

***Correspondence:**
Klaus Wimmers
wimmers@fbn-dummerstorf.de

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *15 November 2018*
Accepted: *01 April 2019*
Published: *17 April 2019*

Citation:
Ponsuksili S, Trakooljul N, Hadlich F, Methling K, Lalk M, Murani E and Wimmers K (2019) Genetic Regulation of Liver Metabolites and Transcripts Linking to Biochemical-Clinical Parameters. Front. Genet. 10:348. doi: 10.3389/fgene.2019.00348

INTRODUCTION

Metabolites are substrates or products of metabolism. As one of the main "omics-" technologies, metabolomics can bridge the phenotype–genotype gap due to the close association of metabolites to cellular biochemical processes (Cascante and Marin, 2008). The metabolome represents the final "omics-" level in the genotype–phenotype map and reflects changes in phenotype and function, whereas the transcriptome and proteome act as mediators of flow (Ryan and Robards, 2006). High-performance metabolic profiling is a high-throughput analysis suitable for routine measurement of endogenous metabolites and metabolic signatures related to health issues (Johnson et al., 2010). Recent advances in bio-analytical technologies allow genome-wide association studies with metabolomics (mGWAS) based on the assumption that the biochemical function of a gene variant is reflected by varied metabolite levels, which are substrates, products, or ligands of that gene product (Adamski and Suhre, 2013).

Association of a single nucleotide polymorphism (SNP) with a metabolic trait indicates that the metabolic phenotype is either a cause or consequence of the metabolic state. Accordingly, it allows generation of biological hypotheses about the role of that metabolite for organismal phenotype (Kathiresan et al., 2009; Franke et al., 2010). Several studies have reported metabolic quantitative trait loci (mQTL) or mGWAS for serum metabolite concentrations in humans (Gieger et al., 2008; Illig et al., 2010; Nicholson et al., 2011). Genetic influences on blood metabolites in healthy humans can be detected by combining genetic variants and metabolic traits (Shin et al., 2014; Draisma et al., 2015).

The regulatory mechanisms between transcript and metabolite levels are still not well understood. Thus, integrating transcriptomics and metabolomics can elucidate the relationship between genes and their transcripts, metabolites, and outcome levels in cells, as reported in microbial, plant, and animal systems (Hoefgen and Nikiforova, 2008; Yang et al., 2009; Yabushita et al., 2013). Expression quantitative trait loci (eQTL) studies are a powerful functional genomics tool, revealing genetic loci that affect RNA transcription levels. eQTL studies facilitate uncovering biological mechanisms that mediate gene regulation and building complex molecular networks for metabolic, biochemical-clinical, and hematological traits (Ponsuksili et al., 2011, 2012, 2016). eQTL studies suggest the potential value of complementary association studies with other molecular traits, such as endocrine or metabolic phenotypes (Ponsuksili et al., 2012; Ghazalpour et al., 2014).

Given the central role of the liver in metabolic and immune functions, we hypothesized that variation of traits related to metabolic state and performance are largely reflected by metabolites and transcripts of hepatic metabolic pathways. Herein, we characterized the genetic landscape of porcine liver metabolites and we linked hepatic metabolite profiles and transcriptomes as well as plasma biochemical-clinical traits in pigs. Analyses of trait-correlated hepatic metabolites and mQTL, together with our previous eQTL results, provide a fine map of loci controlling metabolic profiles. Because pigs are valuable models, this knowledge provides a rational basis not only for understanding pig physiology, but also for human medical research.

MATERIALS AND METHODS

Animals and Sample Collection

Pigs from a German Landrace herd were reared, performance tested, sampled, and used for genome-wide association studies of liver metabolites. Animal care and tissue collection procedures were approved by the Animal Care Committee of the Leibniz Institute for Farm Animal Biology and carried out in accordance with the approved guidelines for safeguarding good scientific practice at the institutions of the Leibniz Association. Measures have been taken to minimize pain and discomfort in line with the guidelines laid down in the Council Directive 86/609/EEC of 24 November 1986. Veterinary inspection of live pigs and their carcasses and organs after slaughter confirmed a lack of any impairments, disease symptoms, or pathological signs to avoid any bias of blood phenotypes. Liver and blood samples were collected from pigs at an average age of 170 days at the experimental slaughter facility of the Leibniz Institute for Farm Animal Biology, between 8.00 and 10.00 in the morning.

Plasma Analyte Measurement

Plasma cortisol concentrations (total) were determined using commercially available enzyme-linked immunosorbent assays (DRG, Marburg, Germany), performed in duplicate according to the manufacturer's protocol. Biochemical-clinical parameters of blood samples were determined using an automated analyser device (Fuji DriChem 4000i, FujiFilm, Minato, Japan) including albumin (ALB), ammonia nitrogen (NH_3), blood urea nitrogen (BUN), glucose (GLU), inorganic phosphorus (IP), and creatinine (CREA).

Metabolic Profiling

A total of 350 individual porcine livers from the same animals used for biochemical-clinical blood plasma analyses were subjected to metabolite profiling. Liver was ground under liquid nitrogen into a homogeneous mixture before being divided for extraction using two-step extraction methods from Wu et al. (2008). We homogenized 50 mg frozen liver powder in 4 mL/g cold methanol and 0.85 mL/g cold water in homogenization tubes containing ceramic beads. Three internal standards were used, including 1 mM ribitol and 0.2 mM palmitic acid-d31 for GC-MS, 250 μM camphorsulphonic acid for LC-MS. Homogenates were transferred to 1.8-mL glass vials and mixed with 2 mL/g chloroform. Samples were vortexed for 60 s, left on ice for 10 min to partition, and centrifuged. Polar and non-polar layers were removed and dried, although we only concentrated on polar phase metabolites in this study. We analyzed samples using non-targeted metabolic profiling instrumentation combining two platforms, GC-MS and HPLC-MS. Both methods represent relative metabolite amount per liver sample (25 mg wet weight of liver per sample). After extraction, samples were split for GC-MS and HPLC-MS analysis, frozen,

and lyophilized. Details of GC-MS and HPLC-MS setups are done according to manufacturer's instructions. In brief, lyophilized samples were derivatized and centrifuged. The supernatant was transferred to a new vial before injection for GC-MS. Qualitative and quantitative analyses were performed using ChromaTOF software v4.50.8.0 (LECO Corporation, United States). HPLC-MS analysis was performed using an Agilent 1100 series liquid chromatographic system (micrOTOF, Bruker Daltonik GmbH, Germany). For analysis, lyophilized liver extracts and blank samples were dissolved in 100 μL water and centrifuged. For chromatographic separation, 5 μL of each sample were injected into a Synergi 2.5 μm Fusion RP column attached to a guard column of the same material (**Supplementary Methods, Data Sheet 1**). Metabolite identification was verified and analysis using the software DataAnalysis v4.0 and QuantAnalysis v2.0 (Bruker Daltonik GmbH, Germany).

SNP Genotype and mRNA Expression Profile Data

Single nucleotide polymorphism genotyping and mRNA hepatic expression profiling was performed using samples of identical animals as for biochemical-clinical blood plasma analyses and liver metabolite profiling. In brief, genotyping was performed using the PorcineSNP60 BeadChip (Illumina Inc., San Diego, CA, United States) per the manufacturer's SNP Infinium HD assay protocol. Samples with call rates of <99%, markers with low minor-allele frequency (<5%), and markers that strongly deviated from Hardy–Weinberg equilibrium ($p < 0.0001$) were excluded. The average call rate for all samples was 99.8% ± 0.2 after filtering.

Total RNA was isolated from liver and amplified using an Ambion WT Expression kit (Affymetrix, Thermo Fisher Scientific, Waltham, MA, United States). Subsequently, cDNA was fragmented, labeled, and hybridized to the microarray using Affymetrix standard protocols. Affymetrix Porcine Snowball microarrays containing 47,880 probesets were used to determine expression profiles. Affymetrix Expression Console software was used for robust multichip average normalization and gene detection by applying detection above background algorithm. Expression data are available in the Gene Expression Omnibus public repository (GEO accession number GSE83932: GSM2221843-GSM2222139). Further filtering was done by excluding transcripts with low signals and probes that were present in <80% of samples. In total, 24,904 probes passed quality filtering and were used for further analyses. Both mRNA and SNPs were mapped to the porcine reference genome using Sscrofa 10.2 (Ensembl downloaded from NCBI[1]).

Data Pre-processing and Statistical Analysis

After quality control and filtering for metabolites of low concentrations and samples with low concentrations of analytes as well as outlier animals, 74 out of 90 metabolites from 343 individuals were further analyzed. Z-score for each metabolite was calculated as: (relative metabolite level in the samples – mean of metabolite level in the samples)/*SD* of metabolite levels in the samples. Metabolite data were further pre-processed to account for systemic effects. Mixed-model analyses of variance using JMP Genomics (SAS Institute, Cary, NC, United States) were used to adjust for fixed and random effects. The genetic similarity matrix between individuals was first computed as identity-by-descent of each pair for the k-matrix and considered as a random effect. For control of population stratification, top principal components (PCs) explaining >1% of variation were considered as covariates. In total, 15 PCs were included as covariates. Gender was used as a fixed effect, batches of metabolite measurement were used as a random effect, and carcass weight was considered as a covariate. Residuals were retained for further analysis.

Metabolite QTL (mQTL) analyses were conducted using the R-package Matrix eQTL (Shabalin, 2012). Matrix eQTL tests for association between each SNP and residual metabolite levels by modeling the additive effects of genotypes in a least squares model (Shabalin, 2012). It performs a separate test for each metabolite–SNP pair and corrects for multiple comparisons by calculating the false discovery rate (FDR).

Residuals of mRNA transcript abundances, after correction for fixed effects (gender), random effects (genetic similarity matrix), and covariates (17 top PCs explaining >1% variation; carcass weight), were used to analyze eQTL by the same process used for mQTL in our previous study (Ponsuksili et al., 2016). We defined an eQTL as cis if an associated SNP was located within an area <1 Mb from the probeset/gene.

Residuals of mRNA and metabolite levels were used for pleiotropic association analyses to identify common regions. Multivariate analysis of variance (MANOVA) between residuals of metabolite and mRNA transcript levels and genetic marker data was used to analyze pleiotropic associations.

Weighted Gene Co-expression Network Analysis (WGCNA)

Residuals of mRNA and metabolite levels were also used to construct co-expression/co-abundance networks using the blockwise modules function of the weighted gene co-expression network analysis (WGCNA) package in R (Langfelder and Horvath, 2008; Ponsuksili et al., 2015). Module–trait associations were estimated using the correlation between module eigengene which is the first PC of module of transcripts and of metabolites and plasma biomarkers. Correlations of metabolites with biochemical-clinical traits and mRNA transcript levels were estimated using Spearman coefficients and corrected for multiple comparisons by calculating FDR. Networks of genes and metabolites were visualized with Metscape 2[2] (Karnovsky et al., 2012).

NADP/NADPH Measurements

In order to validate the correlations found between transcripts and metabolites of the pentose phosphate pathway (PPP), NADPH concentration and NADP/NADPH ratio were

[1]http://www.ncbi.nlm.nih.gov

[2]http://cytoscape.org

measured from liver tissues of a random subset of animals (n = 27) using a NADP/NADPH assay kit (Abcam, Cambridge, United Kingdom) according to manufacturer's instructions. Briefly, 50 mg of liver were washed and homogenized with extraction buffer and then centrifuged to isolate the NADPH/NADP+-containing supernatant. Supernatant was filtered through a 10-kD spin column to remove enzymes that may rapidly consume NADPH. An aliquot of supernatant was heated at 60°C for 30 min to decompose NADP+, cooled on ice, and spun quickly to remove the precipitate. Another aliquot of supernatant was not heated. Both aliquots were reacted with NADP+ cycling buffer and enzyme mix for 5 min at room temperature to convert NADP+ to NADPH. Solutions were then incubated with NADPH developer and absorbance was measured at 450 nm after 1, 2, or 3 h. Amount of NADPH (heated sample) and total NADP+ and NADPH (unheated sample) were quantified from a NADPH standard curve. In the same samples, expression levels of *HDAC4* and *G6PD*, which is the first limiting enzyme of PPP, were determined by qPCR validation. Three reference genes (*RPL32*, *RPS11*, and *ACTB*) were used, and all measurements were performed in duplicate.

Cell Culture and siRNA Transfection

Human HepG2 cells were cultured in DMEM containing L-glutamine, 4.5 g/L D-glucose, and sodium pyruvate (Life Technologies) supplemented with 10% FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin; the medium was refreshed every 2 days. Cell incubation was performed at 37°C in a humidified 5% CO_2 atmosphere. Synthetic siRNAs were pre-designed by Qiagen. A total of four pre-designed siRNAs (Qiagen) per gene were first tested. The most two effective siRNA for *HDAC4* were used (Hs_HDAC4_3 FlexiTube siRNA and Hs_HDAC4_7 FlexiTube siRNA). The average values of negative non-silencing control siRNA (AllStars Negative Control siRNA, Qiagen), mock, and untreated were used as control. Transfection of siRNA was carried out using the HiPerFect transfection reagent (Qiagen) at 150 nM final concentration. The complexes were added drop-wise onto the cells, and the plates were then gently swirled to ensure uniform distribution of the transfection complexes. Forty-eight hours after siRNA transfection, cells were rinsed two times with PBS. The transfected cells were harvested for monitoring the effect of gene silencing. Three independent experiments were conducted. We determined the level of knockdown of *HDAC4* and *G6DP* using quantitative PCR (qPCR) (Roche, Germany) and normalized data using ß-actin as an internal control. All statistical analyses were performed using two-tailed Student's t-tests.

RESULTS

The links between plasma biomarkers, hepatic metabolites, transcripts, and genotypes obtained from some 300 animals reared and performance tested under standardized conditions were analyzed and integrated in this study. Therefore, networks were obtained between metabolites and transcripts; both, from single and weighted correlation network analysis (WGCNA) of transcripts and metabolites (Langfelder and Horvath, 2008; Ponsuksili et al., 2015). Genetic regulation of metabolites (mQTL) was identified and integrated with a genome wide association study of transcripts levels (eQTL) (Ponsuksili et al., 2016). Pleiotropic effects of genetic regions that concertedly regulate transcripts and metabolites were considered. Finally, mQTL, eQTL, and phenotype of blood biochemical-clinical were integrated. The experimental flow is outlined in **Figure 1**.

Metabolite Profiling

In total, we examined 74 liver metabolites of 343 pigs using mass spectrometry and found significant correlations between metabolites (**Figure 2**). Most metabolites in the same molecule class, such as amino acids or nucleotides, clustered together. Metabolite set enrichment analysis of 74 metabolites identified the highest enrichment for protein biosynthesis (16/19), followed by gluconeogenesis (14/27) and glycolysis (12/21) (**Figure 3**). Pathways which reached FDR < 5% are listed in **Supplementary Table S1** together with metabolites within these pathways.

Biochemical-Clinical Traits and Metabolites

Liver metabolites were used for correlation analysis with approved plasma biochemical-clinical biomarkers (ALB; NH_3; BUN; GLU; IP; CREA; and cortisol levels). Three main classes of metabolites with the same profile were identified using WGCNA including carbohydrates, amino acids, and nucleotides. Plasma GLU was found highly positively correlated with eigengene vector of the carbohydrate module and negative correlated with amino acid module (**Figure 4A**).

At a significance level of FDR < 5%, we identified 197 pairs of correlated hepatic metabolites and plasma biomarkers (**Supplementary Table S2**). Correlations between metabolites and biochemical-clinical traits ranged from 0.12 to 0.78. Overall, there was divergent correlation of biochemical-clinical biomarkers with carbohydrate- or amino acid-related metabolites on the one hand and nucleotide metabolism on the other hand. In particular, urea in liver was significantly correlated with BUN in plasma (r = 0.78; $p < 10^{-16}$), as was liver D-glucose with plasma GLU (r = 0.45; $p < 10^{-16}$). Significantly negative correlations were found between plasma GLU and cytidine monophosphate (CMP), inosine monophosphate (IMP), and guanosine monophosphate (GMP) (r = 0.56–0.29; $p < 10^{-8}$). Plasma CREA was significantly negatively correlated with many amino acids, including L-isoleucine, L-tyrosine, L-leucine, L-threonine, L-valine, and L-asparagine (r = 0.13–0.17; $p < 10^{-3}$). In addition, liver 4-hydroxyl-L-proline was significantly positively correlated with plasma CREA (r = 0.32; $p = 1 \times 10^{-9}$). Interestingly, plasma cortisol was significantly negatively correlated with liver D-glucose (r = 0.29; $p = 1 \times 10^{-7}$) and lactate (r = 0.28; $p = 1 \times 10^{-7}$) and positively correlated with IMP (r = 0.35; $p = 9.9 \times 10^{-11}$) and CMP (r = 0.30; $p = 2.3 \times 10^{-8}$).

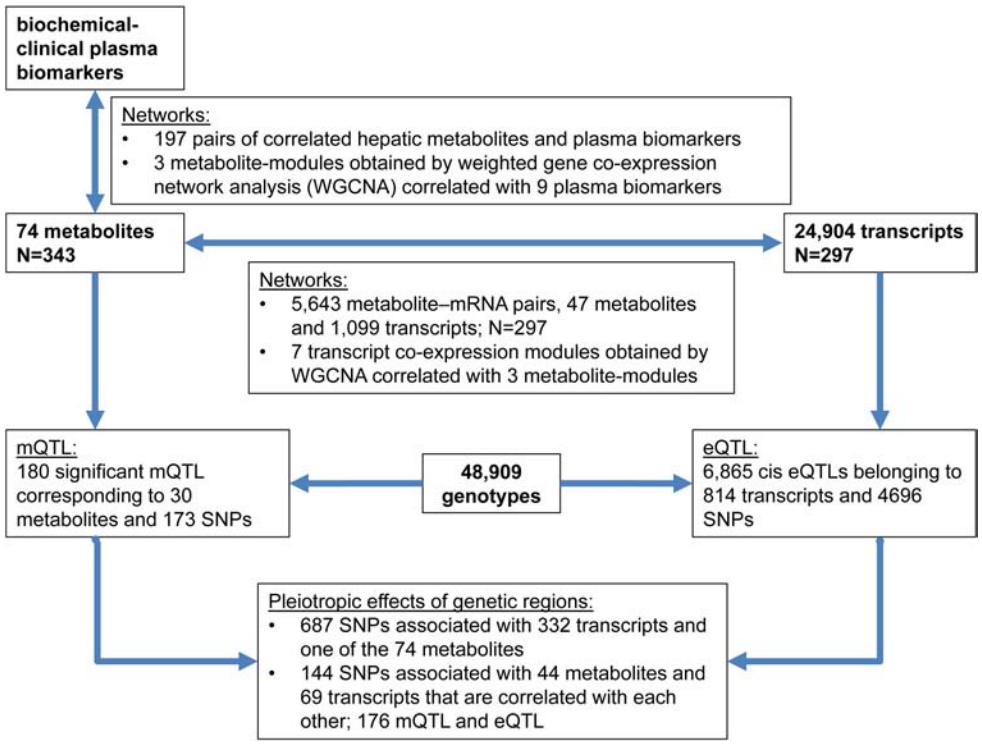


FIGURE 1 | Outline of experimental flow and summary of main results.

Transcripts and Metabolites

Weighted gene co-expression network analysis was performed using the transcriptome data from 24,904 liver transcripts. Seven modules of co-expressed transcripts were highly correlated with metabolite classes, as shown in **Figure 4B**. The co-expressed transcripts in each module were assigned to three top canonical pathways (**Figure 4B**). The amino acid module was significantly positively correlated with immune or acute phase response signaling, whereas the carbohydrate module was highly enriched in cholesterol biosynthesis. We explored transcriptional changes not only in terms of gene co-expression networks but also at the level of individual genes. Pair-wise correlations between the abundance of 24,904 liver transcripts and 74 metabolites in 297 individuals revealed 5643 metabolite–mRNA pairs with correlation coefficients of $r > |0.40|$, corresponding to $p < 3.4 \times 10^{-12}$ and FDR $< 1.1 \times 10^{-9}$. This covered 47 metabolites and 1099 annotated transcripts (1449 probesets). **Supplementary Table S3** shows the top 20 transcripts that are correlated with the individual metabolites. A network-based approach was used to demonstrate the top relationship among transcripts and metabolites (**Figure 5** and **Supplementary Table S3**). The most dominant pathways in these top pairs of metabolites and mRNA were related to PPP (D-ribose 5-phosphate, amino-D-fructose 6-phoshate, D-sedoheptulose 7-phosphate, D-erythrose 4-phosphate), purine (GMP, GDP, IMP), and pyrimidine metabolism (UMP and CMP).

Highly negative correlation was found between *LOC100738008* (thyroid hormone-inducible hepatic protein, *THRSP*) with IMP and CMP ($r = -0.75$ $p < 10^{-16}$) followed by *HDAC4* with D-erythrose 4-phosphate ($r = -0.69$, $p < 10^{-16}$). Expression levels of *HDAC4* were highly positively correlated with CMP, IMP, and UMP. In contrast, *HDAC4* levels were strongly negatively correlated with metabolites in carbohydrate metabolism, particularly PPP metabolites, including D-fructose, D-glucose, glucose 6-phosphate, D-erythrose 4-phosphate, fructose 6-phosphate, fumaric acid, L-lactic acid, malate, D-ribose 5-phosphate, D-sedoheptulose 7-phosphate, and succinic acid. In addition, strong positive correlation was found between CMP and *NMRAL1* ($r = 0.72$; $p < 10^{-16}$). Furthermore, transcript levels of *ARG2*, followed by *SLC22A7* (organic anion transporter), *XRCC6BP1*, *SLC38A1*, and *SLC7A2*, were highly correlated with most amino acids.

NADP/NADPH Measurements

Because PPP was dominantly linked with *HDAC4*, we measured NADPH concentration and the ratio of NADP/NADPH, i.e., the main products of PPP, as well as expression levels of *HDAC4* and *G6PD*, the key enzyme of PPP, in order to provide experimental evidence of the link of transcripts and PPP activity. Using qPCR, we found significant correlation between NADPH concentration and expression levels of *HDAC4* and *G6PD*. We confirmed expression levels of *HDAC4* obtained from the microarray by qPCR ($r = 0.93$; $p < 0.0001$) while *G6PD* was not available on the Affymetrix chip. Expression levels of *HDAC4* were positively correlated with NADP/NADPH ($r = 0.78$; $p < 0.0001$) and

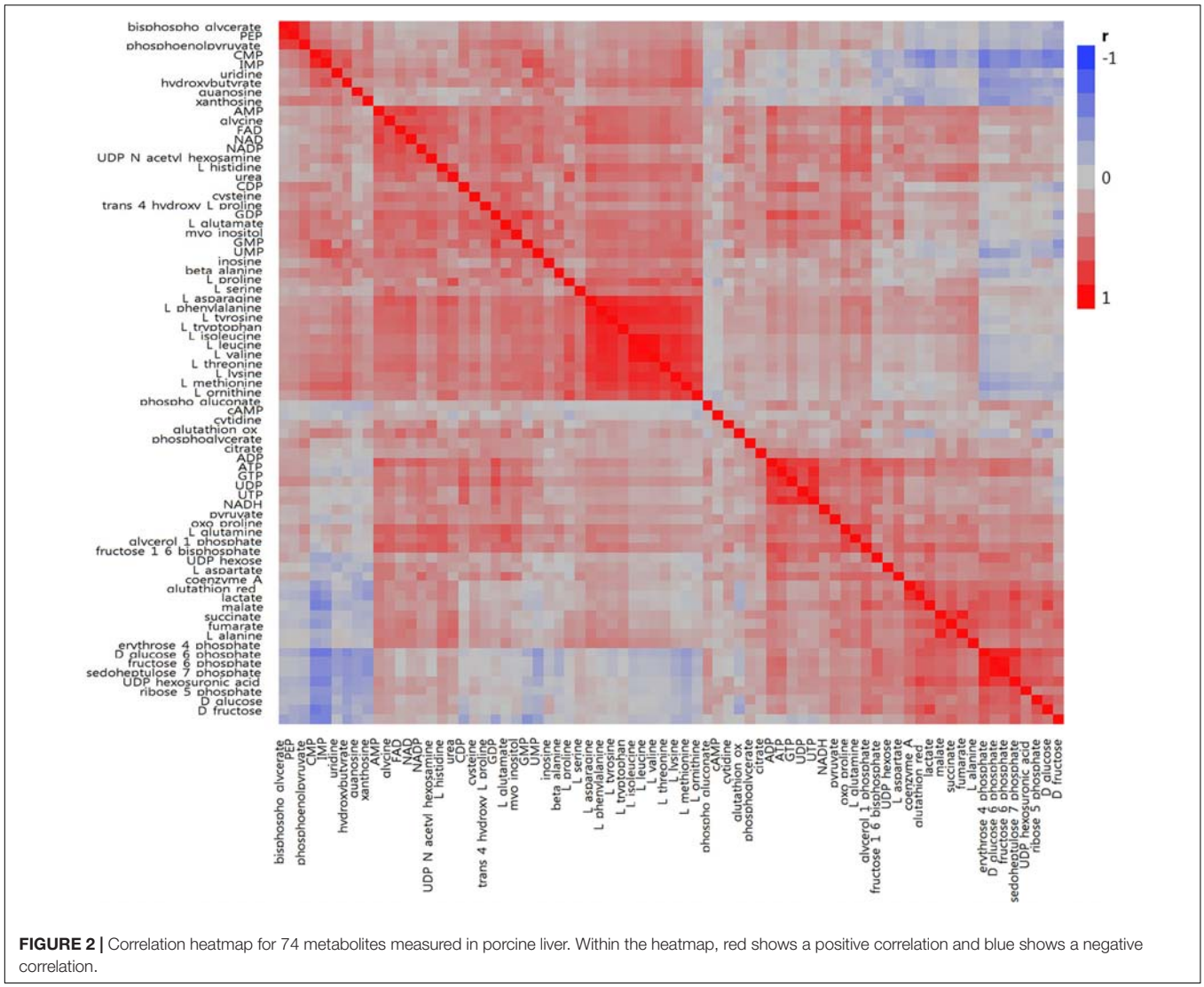


FIGURE 2 | Correlation heatmap for 74 metabolites measured in porcine liver. Within the heatmap, red shows a positive correlation and blue shows a negative correlation.

negatively correlated with NADPH concentration ($r = -0.71$; $p < 0.0001$). *G6PD* had a significant negative correlation with *HDAC4* ($r = -0.44$; $p = 0.02$) but positive correlation with NADPH concentration ($r = 0.61$ and $p = 0.0007$) and negative correlation with NADP/NADPH ($r = -0.47$; $p = 0.012$). *G6PD* expression also was correlated with PPP metabolites, including erythrose 4-phosphate, sedoheptulose 7-phosphate, D-glucose 6-phosphate, and fructose 6-phosphate ($r = 0.58$–0.63; $p = 0.0012$–0.0004).

HDAC4 Knockdown and *G6PD* Expression

To further experimentally elucidate the link of *HDAC4* and *G6PD* expression, RNAi was used to knockdown *HDAC4* expression *in vitro* in the Human HepG2 cells line. Subsequently, relative expression of *G6PD* was measured using qPCR. siRNA targeting *HDAC4* inhibited its expression to 70–80% relative to control cells ($p < 0.004$). At the same time, *G6PD* showed increased expression levels to 120–130% compare to control ($p < 0.003$) leading to pronounced differential expression between *HDAC4* and *G6PD* ($p = 0.0002$) (**Figure 6**).

Genome-Wide Association of Metabolites (mQTL)

A genome-wide association study covering 48,909 SNP genotypes and 74 metabolites revealed 180 significant mQTL that corresponded to 30 metabolites and 173 SNPs at a threshold of $-\log_{10} > 4$ (**Supplementary Table S4**). **Table 1** lists top 10 associations. Only hydroxy-L-proline reached the significance threshold of FDR < 5% while other three metabolites (citrate, cysteine, and beta-alanine) showed suggestive mQTL at FDR ≤ 10%. Percent phenotypic variance explained by peak markers for these four metabolites was 6.7–9.4%. **Figure 7** shows associations of these four metabolites across different pig chromosomes. The strongest association was for trans-4-hydroxy-L-proline with SNPs at 39.9 Mb on chromosome 6 ($p = 6 \times 10^{-9}$) (**Table 1** and **Figure 7A**). Markers at position 53 Mb of chromosome 18 showed significant association with

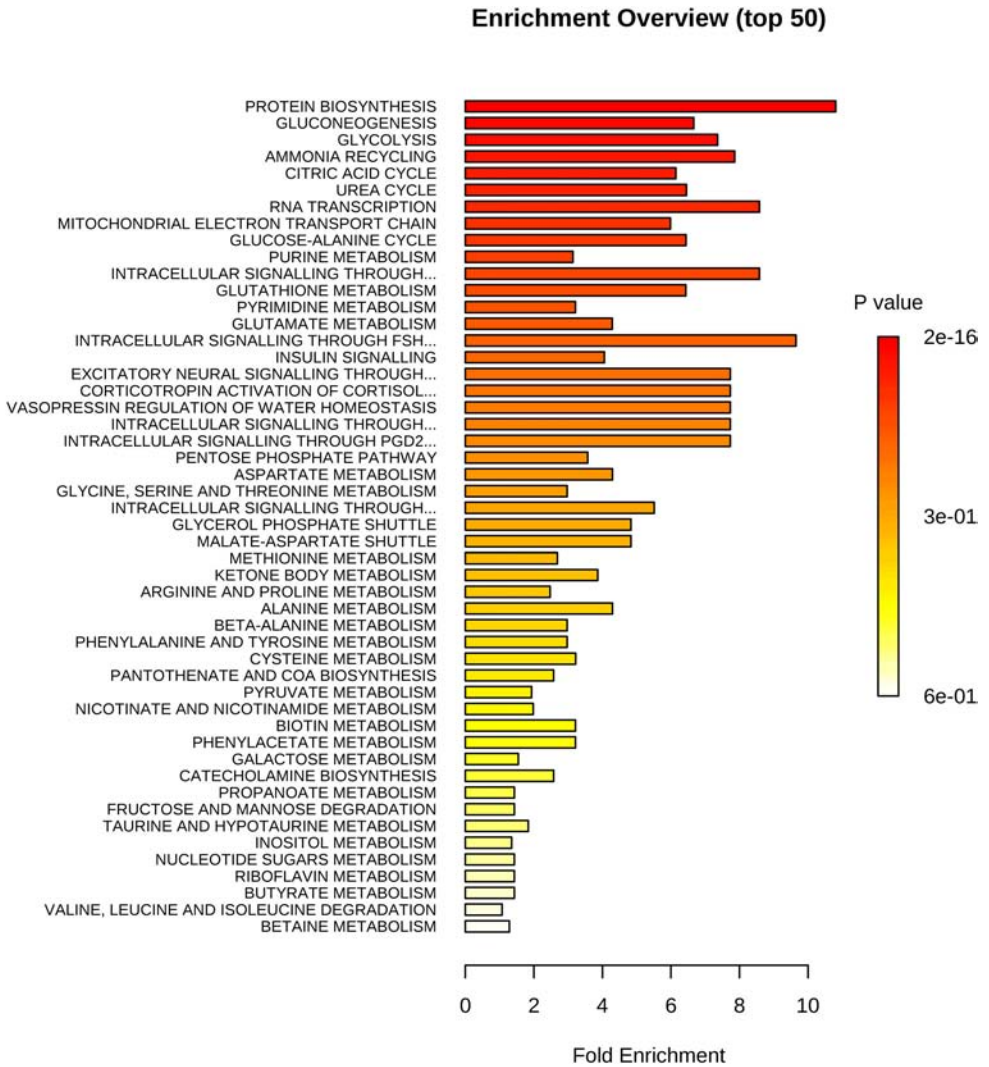


FIGURE 3 | Enrichment analysis of 74 metabolites. Highest enrichment was found for protein biosynthesis (16/19) followed by gluconeogenesis (14/27) and glycolysis (12/21).

beta-alanine (**Figure 7B**). For citric acid (**Figure 7C**) and cysteine (**Figure 7D**), significant markers were mapped at various regions in the genome.

mQTL, eQTL, and Transcript Correlated Metabolites

Metabolic QTL regions contain numerous positional candidate genes, depending upon the level of linkage disequilibrium. To support and narrow down the number of candidate genes in regions, we integrated our previous eQTL data from the same pigs (Ponsuksili et al., 2016). Many SNPs associated with metabolites were also associated with transcripts. In our previous study, 6865 eQTLs were identified as cis, belonging to 1028 probesets (814 annotated transcripts) at FDR < 5% ($p < 10^{-7}$). Further, 687 SNPs that were associated with mRNA transcripts (332 probesets) were associated with one of the 74 metabolites.

In addition, we considered only metabolites that significantly correlated with mRNA transcripts at FDR < 5%. In total, 144 SNPs were associated with 44 metabolites and 69 metabolite-correlated transcripts, representing 176 mQTL and eQTL (**Supplementary Table S5**). Nineteen out of these 144 SNPs on *Sus scrofa* chromosome (SSC) 6 associated with trans-4-hydroxy-L-proline ($p < 6.0 \times 10^{-9}$–1.1×10^{-4}). These SNPs were simultaneously associated with transcript levels of *PRODH2* ($p < 4.7 \times 10^{-26}$–4.9×10^{-11}). Moreover, trans-4-hydroxy-L-proline was negatively correlated with *PRODH2* ($r = -0.40$; $p = 1.6 \times 10^{-12}$). Pleiotropic association analyses also showed SNP-directed links between trans-4-hydroxy-L-proline and *PRODH2* with 91 SNPs on SSC 6 (FDR < 5%) (**Figure 8A**).

At 5% FDR, six SNPs at position 53.4–54.9 Mb on SSC 18 were associated with beta-alanine and transcript levels of *IGFBP-3* (**Figure 8B**). The correlation between beta-alanine and transcript levels of *IGFBP-3* was $r = -0.17$ and $p = 2.8 \times 10^{-3}$.

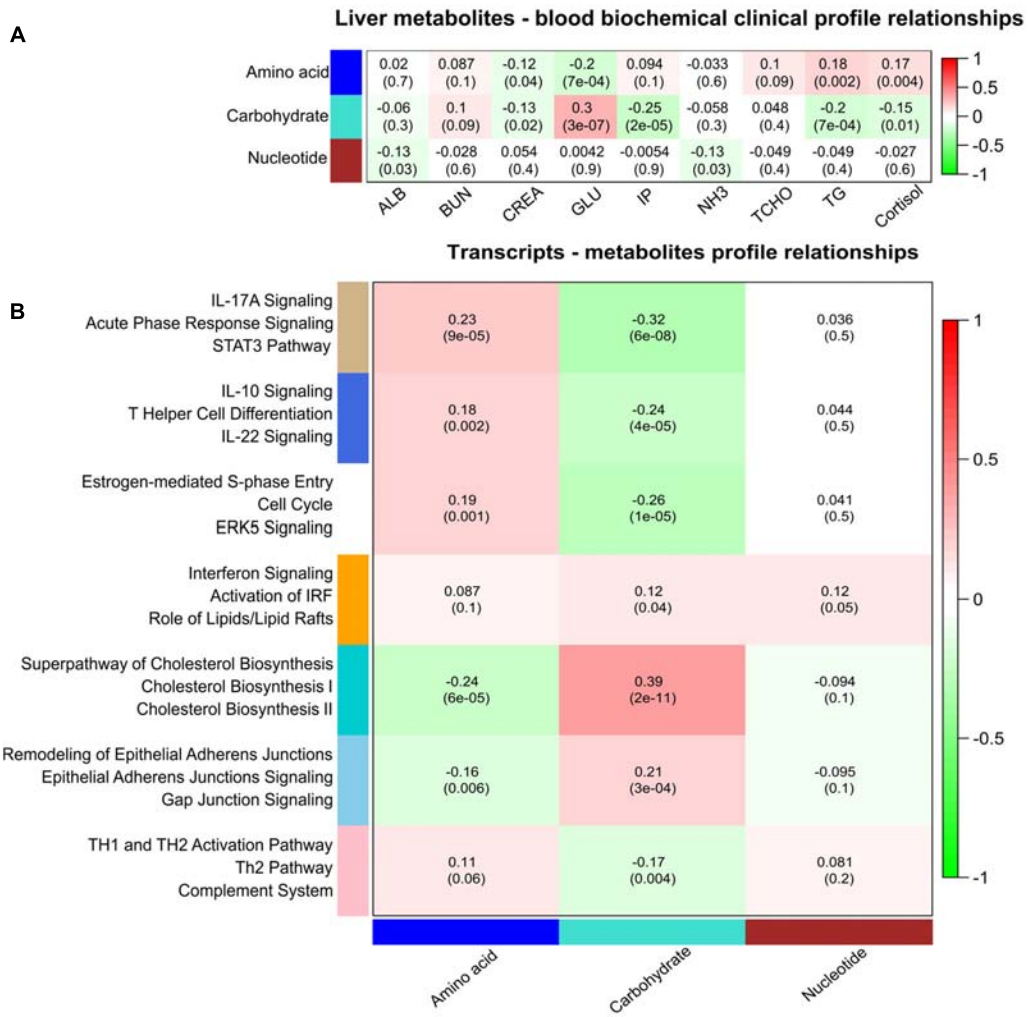


FIGURE 4 | Correlation matrix of module eigengene values obtained for metabolites, transcripts, and plasma biomarkers. Weighted gene co-expression network analysis (WGCNA) groups metabolites and transcripts into modules based on correlated abundances. Each of the modules was labeled with a unique color as an identifier. **(A)** Three modules of metabolites including amino acids, carbohydrates, and nucleotides showing significant correlation with plasma biomarkers. **(B)** Seven modules of co-expressed transcripts showing significant correlation with three modules of metabolites. Within each cell, upper values are correlation coefficients and lower values are the corresponding *p*-values. Canonical pathways related to genes of these seven modules of co-expressed genes are given at the left side.

In other cases, SNPs located on SSC 7 position 20.5 Mb associated with transcript levels of *ALDH5A1* ($p = 5.1 \times 10^{-13}$) were also associated with beta-alanine, although at FDR > 5%. The correlation between *ALDH5A1* and beta-alanine was highly significant ($r = -0.24$; $p = 2.7 \times 10^{-5}$). The highest correlation was found between transcripts levels of *DPYS* and 3-hydroxybutyrate ($r = -0.45$; $p = 2.6 \times 10^{-15}$). Three SNPs located on SSC 4 position 35.6 Mb were associated with *DPYS* ($p = 6.6 \times 10^{-11}$) and, at a lower significance level, with 3-hydroxybutyrate ($p = 1.9 \times 10^{-3}$). As shown in **Figure 7C**, significant markers associated with citrate mapped to various regions in the genome. By combining eQTL, mQTL, and the correlation of corresponding mRNAs and metabolites, we found two interesting candidate genes in peak regions for citrate: *STAB2* on SSC 5 position 84.3 Mb and *MFHAS1* on SSC 15 position 63.7 Mb. Ten SNPs on SSC 15 position 63.7 Mb were associated with both *MFHAS1* ($p = 8.2 \times 10^{-12}$) and citrate ($p = 3.4 \times 10^{-4}$). Eight significant markers associated with *STAB2* ($p = 1.1 \times 10^{-7}$–1.1×10^{-6}) were also associated not only with citrate but also with malate, succinate, pyruvate, and D-fructose ($p = 8.9 \times 10^{-3}$–4.4×10^{-4}). These metabolites, which mostly belong to the citric acid cycle, were also negatively correlated with *STAB2* ($r = 0.21$–0.31; $p = 2.4 \times 10^{-4}$–5.3×10^{-8}). Pleiotropic association analyses of transcript levels of both *STAB2* and *MFHAS1* and the metabolites of citrate, malate, succinate, pyruvate, and D-fructose showed 47 markers located on SSC 5, with 15 reaching a significance threshold of 5% FDR (**Figure 8C**). Another interesting transcript was *RBBP9*, which was negatively correlated with ribose 5-phosphate ($r = 0.16$; $p = 4.5 \times 10^{-3}$) and D-glucose 6-phosphate ($r = 0.30$; $p = 2.9 \times 10^{-7}$). Transcript levels of *RBBP9* were associated with 6 SNPs that were also associated with both ribose 5-phosphate and D-glucose 6-phosphate.

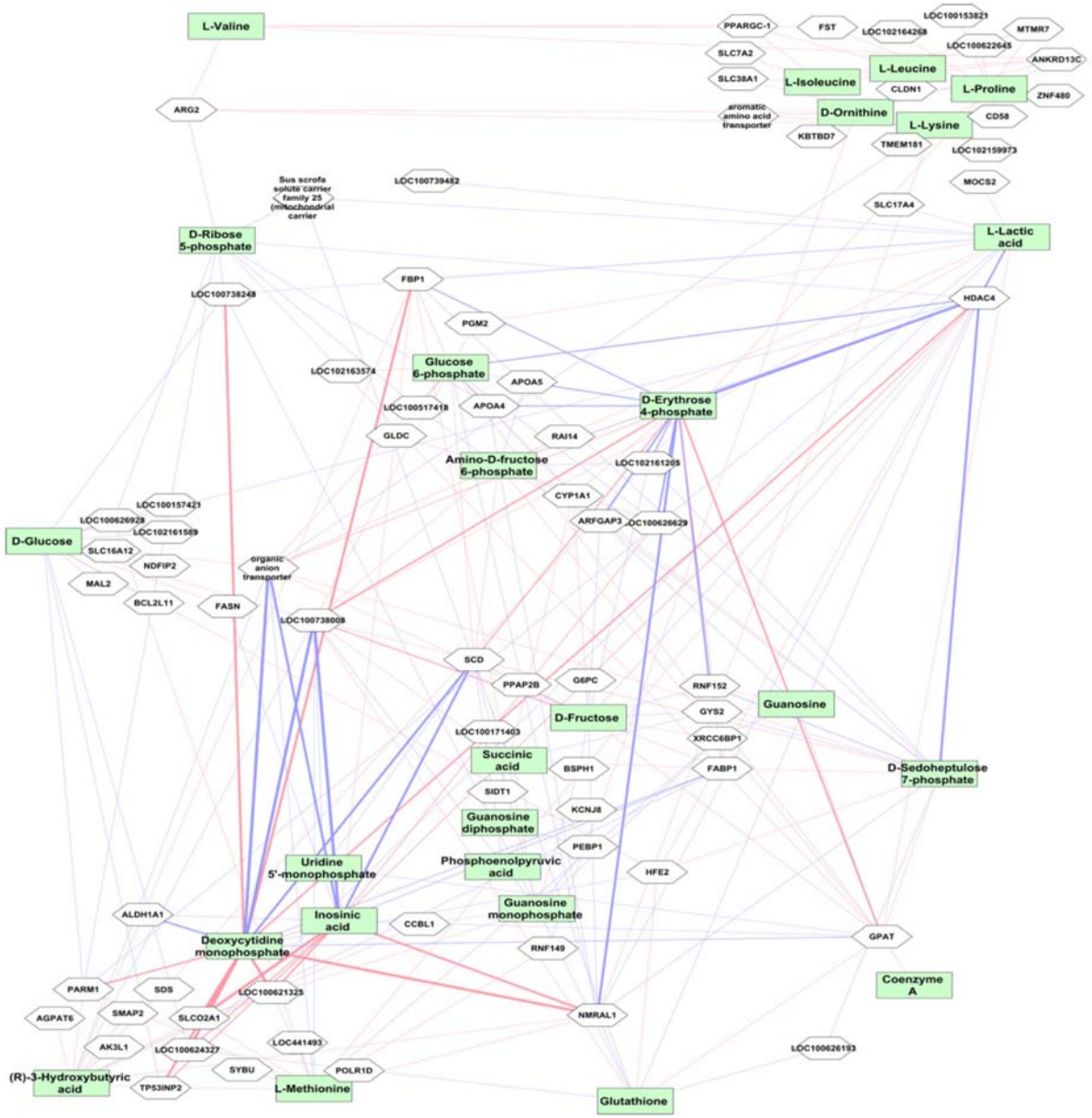


FIGURE 5 | Correlation between mRNA transcript level and metabolites. The top metabolites with strong correlation to transcripts ($r > \pm 0.5$, $p < 10^{-16}$) are shown. The red connections indicate positive correlation and blue color shows negative correlation. The level of correlation is demonstrated by the thickness of the line. The metabolites are shown in green boxes and genes are in white boxes.

DISCUSSION

An improved understanding of non-genetic and genetic regulation of metabolite levels facilitates their interpretation as biomarkers for complex traits related to the metabolic status and in terms of exogenous and endogenous impacts on phenotypes. Moreover, identification of links between genetic polymorphisms and transcript and metabolite levels contributes to the elucidation of biomarkers that are the cause or consequence of changes in metabolic pathways. However, interpretation of mQTL data is demanding due to the fact that many metabolites are involved in various pathways. Here, we investigated a set of metabolites—mostly amino acids, carbohydrates, and nucleotides—in the polar phase of liver extracts.

Correlation Between Biochemical-Clinical Traits, Transcripts, and Metabolites

To understand the relationship between gene expression, metabolite levels, and biochemical-clinical traits using a system genetics approach (Civelek and Lusis, 2014), we

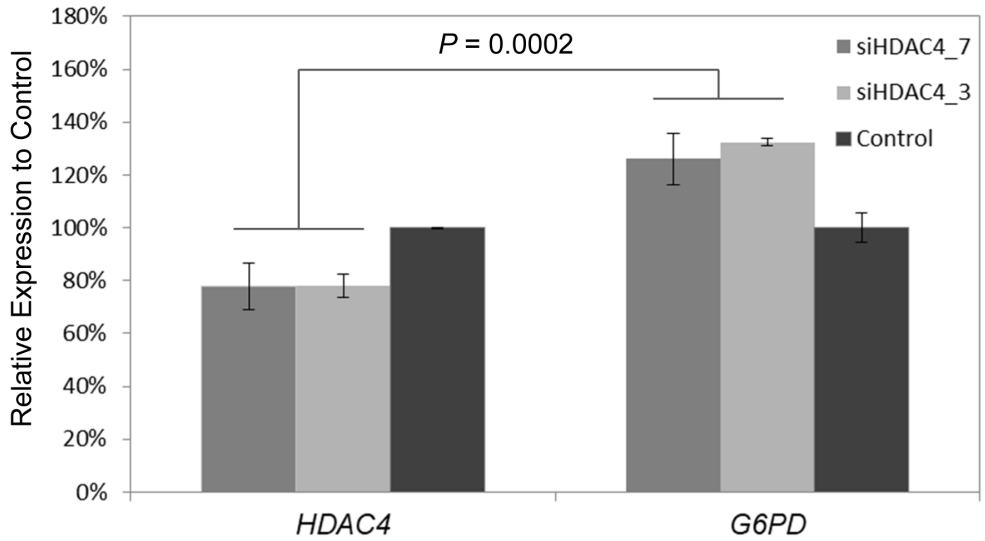


FIGURE 6 | Knockdown of *HDAC4* by RNA interference reveals upregulation of *G6PD*. Two siRNAs siHDAC4_7 and siHADC4_3 were designed to target *HDAC4* and transfected into human HepG2 cells *in vitro*. Relative mRNA expression was measured by qPCR 48 h after transfection. Expression was normalized to ß-actin internal controls. Expression of *G6PD* was significantly increased relative to its expression in control cells and *HDAC4* levels at 48 h post-transfection of siRNA. The data represent means ± *SD* (n = 3).

integrated these data obtained from the same pigs by calculating pair-wise correlations and WGCNA. We found significant intra- and inter-class correlations between metabolites especially amino acids and carbohydrate reflecting shared biochemical pathways or regulatory interactions with immune and cholesterol biosynthesis. The presence of significant correlations between metabolites categorized and biological function of co-expression transcripts presumably reflects either multiple roles of metabolites or interactions between metabolic pathways and immune system. Correlation of metabolites with transcripts can be due to enzymes, receptors, and signals of pathways encoded by corresponding genes or regulatory factors affecting gene expression. We identified many associations that show that the approach is suitable to identify biologically meaningful links between variation at the genome, transcriptome, and metabolome level with clinically relevant phenotypes. Thus, this approach has the potential to detect novel biomarkers while considering the contribution of exogenous and endogenous factors to individual variation.

TABLE 1 | Top 10 mQTL results.

Metabolite	SNP_ID	rs_number	*p*-value	SSC10_2	Base pair
Trans-4-hydroxy-L-proline	MARC0072609		6.04E-09	6	39960478
Citric acid	DRGA0014885		9.31E-08	15	11247606
Cysteine	ALGA0077013	rs80979261	6.46E-07	14	40660903
Beta_alanine	ASGA0085673	rs81327629	1.17E-06	18	53448342
Ornithine	ALGA0110895	rs81339246	2.50E-06	1	
Fumaric acid	ASGA0090623	rs81308979	5.20E-06	X	12973924
Malate	ASGA0090623	rs81308979	6.21E-06	X	12973924
L-Lactic acid	H3GA0017168	rs80890289	6.68E-06	5	99642466
6-Phospho-gluconate	CASI0009941	rs330835858	6.91E-06	16	63944017
NADH	ASGA0086564	rs81309680	7.96E-06	16	

For example, D-erythrose 4-phosphate, fructose 6-phosphate, D-ribose 5-phosphate, and D-sedoheptulose 7-phosphate, which belong to PPP, were highly negatively correlated with transcript levels of *HDAC4*. PPP is one of the fundamental components of cellular carbohydrate metabolism and is especially crucial for cancer cells (Kowalik et al., 2017). We confirmed the association by measuring ratio of NADP/NADPH and concentration of NADPH, for which PPP is the major source, as well as expression of *HDAC4* and *G6PD*. Here we show an association of PPP and *HDAC4* in healthy animals, indicating a possible epigenetic-based link between the histone-modifying *HDAC4* and the PPP-driving *G6PD*. *NMRAL1*, which encodes an NADPH sensor protein, is another transcript negatively correlated with PPP metabolites and contributes to regulation of the oxidative phase of PPP (Barcia-Vieitez and Ramos-Martínez, 2014). In addition, knockdown of *HDAC4* using RNAi was shown to be associated with increasing *G6PD* expression.

The liver plays a central role in processes of glycogenesis, glycogenolysis, and gluconeogenesis and thus glucose homeostasis (Nordlie et al., 1999). Our results demonstrate that plasma GLU is highly positively correlated with liver D-glucose. This also matches the finding that transcript levels of both *HDAC4* and *NMRAL1* are negatively correlated with plasma GLU and liver D-glucose, with the latter two being positively correlated.

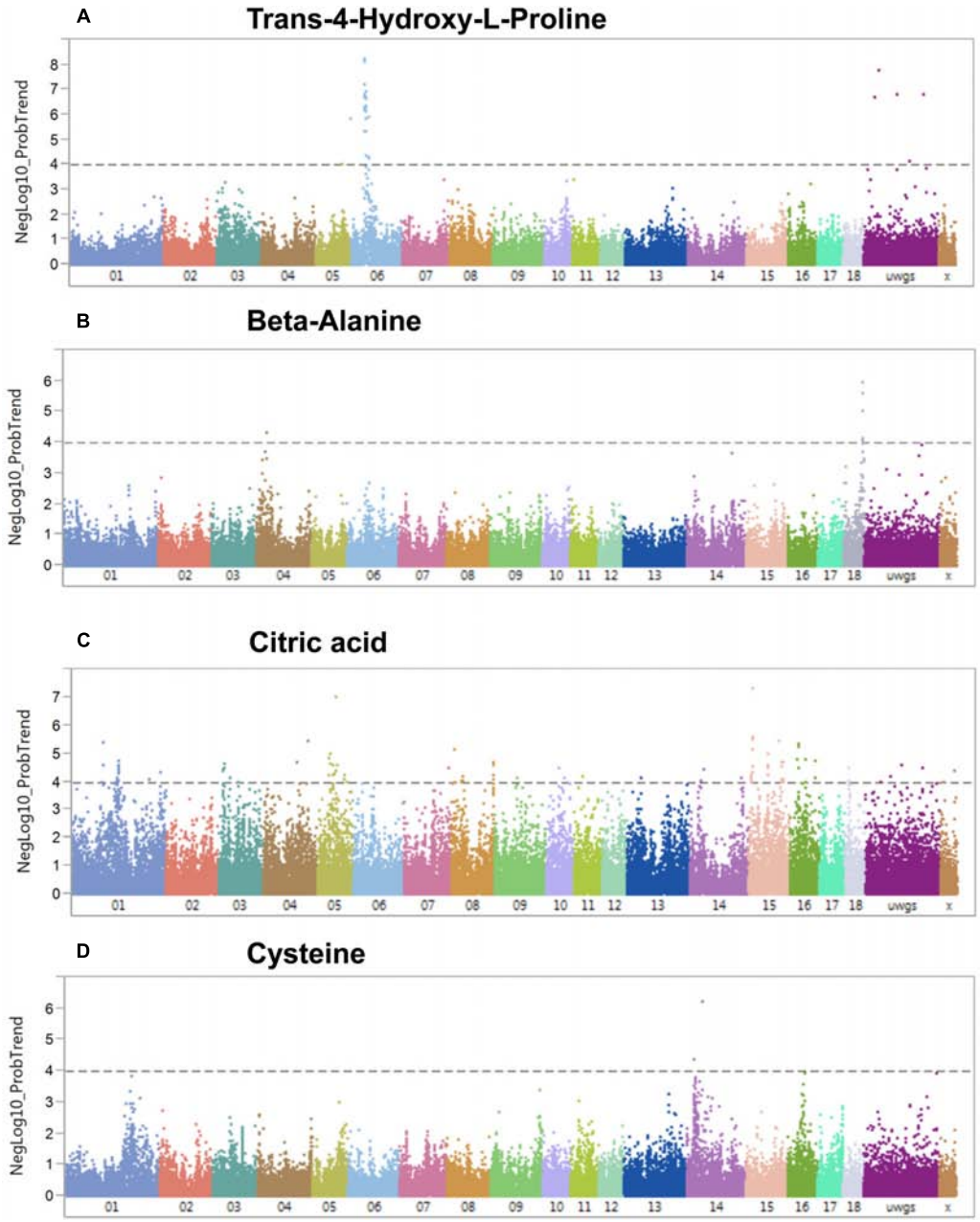


FIGURE 7 | Manhattan plots visualizing genome wide associations of SNPs and metabolites (mQTL). **(A)** Trans-4-hydroxy-L-proline, **(B)** beta-alanine, **(C)** citric acid, and **(D)** cysteine. The dotted line depicts the genome-wide significance thresholds at negative log 10 > 4.

Many transcripts positively correlated with plasma GLU and also correlated with liver metabolites like CMP and IMP, including *THRSP*, *SCD*, and *GPAM*, most of which are involved in lipid metabolism. Thyroid hormone responsive protein (THRSP) is involved in lipogenic processes and is associated with obesity (Ortega et al., 2010) and differential intramuscular fat in cattle (Hudson et al., 2015). Stearoyl-CoA desaturase (*SCD*) is a rate-limiting enzyme in fatty acid biosynthesis and thus a crucial control point of hepatic lipogenesis and lipid oxidation. Glycerol-3-phosphate acyltransferase (*GPAM*) encodes a mitochondrial enzyme that preferentially accepts saturated fatty acids as substrates for glycerolipid synthesis. Together, we show a link between liver metabolites and transcripts involved in lipid metabolism and plasma biochemical-clinical traits.

We found plasma cortisol levels were negatively correlated with liver metabolites that are mostly involved in glucose metabolism. Plasma cortisol levels also positively correlated with liver metabolites like CMP, IMP, and GMP, which

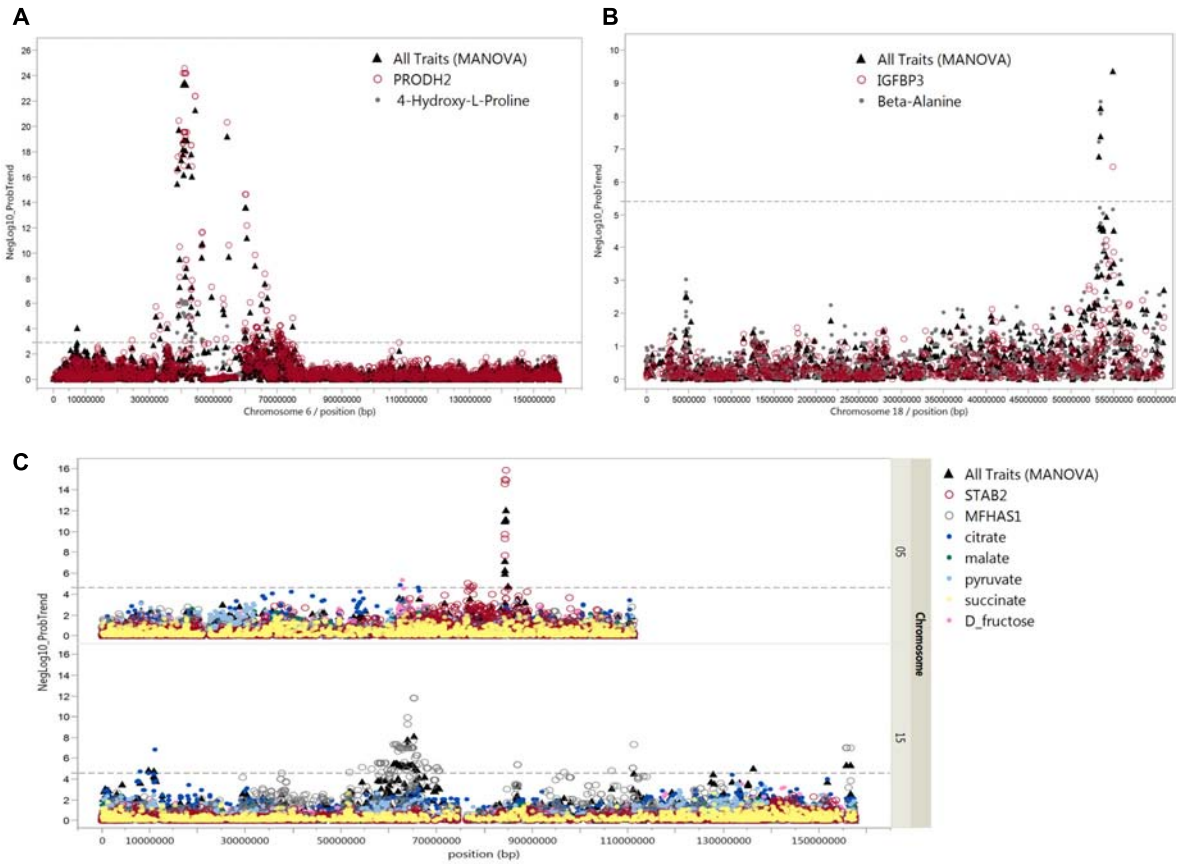


FIGURE 8 | Pleiotropic associations of metabolites and mRNAs. Manhattan plots of pleiotropic associations between metabolites and mRNA expression. The pleiotropic association of these transcripts and metabolites (all traits) was significant (FDR ≤ 5%). **(A)** cis-eQTL of *PRODH2* and mQTL for trans-4-hydroxy-L-proline in close vicinity on chromosome 6, **(B)** cis-eQTL of *IGFBP3* and mQTL for beta-alanine in close vicinity on chromosome 18, and **(C)** cis-eQTL of *STAB2* on chromosome 5 and *MFHAS1* on chromosome 15 with mQTL for citrate, malate, pyruvate, succinate, and D-fructose. The *x*-axis indicates chromosome locations and *y*-axis shows –log10 of the *p*-values of multivariate analysis of variance (MANOVA). The dashed line shows the levels of *p*-values which are significant at 5% FDR.

in turn correlated with transcripts involved in lipid metabolism. This finding confirms our previous study, where we demonstrated these linked biological functions and molecular pathways using an integrative multi-omics approach (Ponsuksili et al., 2012).

Administration of two nucleotides, CMP and UMP, favors the entry of glucose in muscle and maintenance of hepatic glycogen levels during exercise (Gella et al., 2008). Interestingly, we found that cortisol-mediated homeostasis of lipid and carbohydrate metabolism in liver was associated with transcript levels of *CREM*. Abundance of *CREM* transcripts negatively correlated with plasma GLU and liver metabolites of carbohydrate metabolism (D-fructose, D-glucose, ribose 5-phosphate, erythrose 4-phosphate, sedoheptulose 7-phosphate, and lactate) and, at the same time, positively correlated with cortisol levels. *CREM* encodes a transcription factor that binds to cAMP responsive elements to mediate signal transduction during complex processes (Kirchhof et al., 2013; Ella et al., 2014). Previous studies show that *Crem* knock-out mice exhibit less anxious behaviors than wild-type mice (Maldonado et al., 1999). *CREM* is involved in cancer (Passon et al., 2012) and circadian regulation of cholesterol synthesis in the liver (Acimovic et al., 2008). Together, our results link hormone levels in plasma with metabolite and transcripts levels in liver.

ARG2 encodes arginase, which is the enzyme of the final step of the ornithine-urea cycle converting L-arginine to L-ornithine and urea. In the present study, expression of *ARG2* was highly correlated with most amino acids, including L-isoleucine, L-leucine, L-lysine, L-methionine, L-ornithine, L-proline, and L-valine. These amino acids were also negatively correlated with plasma CREA. Transcript levels of *ARG2* also were negatively correlated with plasma CREA and positively correlated with plasma BUN. *Arg2*−/− mice have lower plasma CREA and BUN levels after renal injury (Raup-Konsavage et al., 2017). Our study shows that *ARG2* plays a central role for most amino acid metabolites in liver and is linked to biochemical properties of blood.

Our study highlights the value of integrating data from the same animals from various -omics levels, including transcriptome, metabolome, and biochemical-clinical traits that share biological pathways or functions. We found that epigenetic

modifications mediated by *HDAC4* may play a significant role in PPP. Further, liver metabolites of the nucleotide class linked transcripts involved in lipid metabolism and cortisol. Finally, significant transcripts, such as *ARG2*, linked most amino acids in liver and biochemical-clinical traits, including CREA and BUN.

Comprehensive metabolite screens in the porcine model have identified novel associations among transcript levels, metabolites, and biochemical-clinical traits. Several studies have addressed the genetic regulation of metabolites serving as biomarkers for diseases (Illig et al., 2010; Wang et al., 2011; McMahon et al., 2017; Zhang et al., 2017). However, most studies have measured metabolites in blood serum or urine, while few have focused on genetic regulation of metabolites in other tissues, such as liver or fat (Ghazalpour et al., 2014; Parks et al., 2015). In this study, we integrated genetic-regulated liver metabolites, liver transcripts (mQTL and eQTL), and plasma biochemical-clinical traits. We prioritized genes based on cis-eQTL. For genome-wide significant loci associated with trans-4-hydroxy-L-proline, we identified *PRODH2* as significantly associated with the same SNPs. In addition, we demonstrated that these SNPs show pleiotropic effects by simultaneously affecting trans-4-hydroxy-L-proline and *PRODH2* expression. Further, we identified *PRODH2* as a high-confidence candidate gene within a locus associated with trans-4-hydroxy-L-proline, which in turn strongly correlated with plasma CREA. Trans-4-hydroxy-L-proline is metabolized by the liver and kidneys (Knight et al., 2009). Proline dehydrogenase 2 (*PRODH2*) catalyzes the first enzymatic step in the hydroxyproline catabolic pathway in liver and kidney mitochondria. In addition, *PRODH2* is reported as a molecular target for treating primary hyperoxaluria (Summitt et al., 2015). Mutations in *PRODH2* cause human hydroxyprolinemia, which hampers dehydrogenation of hydroxyproline to delta1-pyroline-3-hydroxy-5-carboxylic acid (Staufner et al., 2016).

In this study, we found a highly negative correlation between *DPYS* and 3-hydroxybutyric acid and identified three SNPs regulating both. Moreover, we found 3-hydroxybutyric acid correlated with cortisol. *DPYS* encodes dihydropyrimidinase, which is the second enzyme of the pyrimidine degradation pathway. The facts that patients with dihydropyrimidinase deficiency show mainly neurological and gastrointestinal abnormalities (van Kuilenburg et al., 2010) and that hydroxybutyric acid passes through the blood–brain barrier into the central nervous system (Sleiman et al., 2016) provide a possible link between *DPYS* and hydroxybutyric acid. Our study provides further evidence for this relationship. However, the link to cortisol as shown here is novel and still unclear.

IGF-binding protein-3 (*IGFBP-3*) is the major carrier protein for IGF-1 and plays a role in cancer, apoptosis, and pathogenesis of ischemia reperfusion after liver injury (Lee et al., 2014; Zhou et al., 2015; Wang et al., 2017). High *IGFBP-3* levels impact myogenesis and enhance muscle protein degradation (Huang et al., 2016). Patients with non-alcoholic steatohepatitis have increased levels of hepatic alanine (Kim et al., 2017). In this study, we found for the first time a link between genetic regulated alanine levels (mQTL) and *IGFBP-3* (cis-eQTL)x.

Genetically regulated metabolites belonging to the citrate cycle (D-fructose, malate, succinate, pyruvate, and citrate) share SNPs that also are associated with transcript levels of *STAB2* and *MFHAS1* (cis-eQTL). The biological function of both transcripts linked via common SNPs and to liver metabolites is still unknown. Here, SNPs located on SSC 17 position 27.4 Mb were associated with transcript levels of *RBBP9* (cis-eQTL) and also with ribose 5-phosphate and glucose 6-phosphate levels, both PPP metabolites. Glucokinase phosphorylates glucose to glucose 6-phosphate in liver as a substrate for several metabolic pathways, including PPP, which is particularly important in rapidly dividing cells like cancer cells for DNA replication. Further, previous studies have reported retinoblastoma binding protein 9 (*RBBP9*) is a tumor-associated protein in pancreatic neoplasia, affecting cell cycle control and contributing to the TGF-β signaling pathway (Acimovic et al., 2008; Vorobiev et al., 2012).

CONCLUSION

In summary, this study is the first to combine metabolomics, transcriptomics, and genome-wide association studies in a porcine model. Our results improve understanding of the genetic regulation of metabolites which link to transcripts and finally biochemical-clinical parameters. Further, high-performance profiling of metabolites as intermediate phenotypes is a potentially powerful approach to uncover how genetic variation affects metabolic and health status. Our results advance knowledge in areas of biomedical and agricultural interest and identify potential correlates of biomarkers, SNPs-metabolites, SNPs-transcripts, and biochemical-clinical traits.

ETHICS STATEMENT

Animal care and tissue collection procedures were approved by the Animal Care Committee of the Leibniz Institute for Farm Animal Biology and carried out in accordance with the approved guidelines for safeguarding good scientific practice at the institutions in the Leibniz Association and the measures were taken to minimize pain and discomfort and accord with the guidelines laid down by the European Communities Council Directive of 24 November 1986 (86/609/EEC).

AUTHOR CONTRIBUTIONS

SP and KW designed the study and interpreted the data. SP performed the statistical and bioinformatic analyses and drafted the manuscript. FH helped in bioinformatics analyses. EM and NT sampled the tissue probes and obtained biochemical-clinical data. KM and ML performed non-targeted

metabolic profiling. FH, NT, EM, KM, ML, and KW critically revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by internal funds of the FBN and received additional funding of the Federal Ministry of Education and Research (BMBF) as part of the PHENOMICS project (Grant No. 0315536F).

ACKNOWLEDGMENTS

The authors thank Joana Bittner, Nicole Gentz, and Annette Jugert for excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00348/full#supplementary-material

REFERENCES

Acimovic, J., Fink, M., Pompon, D., Bjorkhem, I., Hirayama, J., Sassone-Corsi, P., et al. (2008). CREM modulates the circadian expression of CYP51, HMGCR and cholesterogenesis in the liver. *Biochem. Biophys. Res. Commun.* 376, 206–210. doi: 10.1016/j.bbrc.2008.08.126

Adamski, J., and Suhre, K. (2013). Metabolomics platforms for genome wide association studies–linking the genome to the metabolome. *Curr. Opin. Biotechnol.* 24, 39–47. doi: 10.1016/j.copbio.2012.10.003

Barcia-Vieitez, R., and Ramos-Martínez, J. I. (2014). The regulation of the oxidative phase of the pentose phosphate pathway: new answers to old problems. *IUBMB Life* 66, 775–779. doi: 10.1002/iub.1329

Cascante, M., and Marin, S. (2008). Metabolomics and fluxomics approaches. *Essays Biochem.* 45, 67–81. doi: 10.1042/BSE0450067

Civelek, M., and Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* 15, 34–48. doi: 10.1038/nrg3575

Draisma, H. H., Pool, R., Kobl, M., Jansen, R., Petersen, A. K., Vaarhorst, A. A., et al. (2015). Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat. Commun.* 6:7208. doi: 10.1038/ncomms8208

Ella, E., Heim, D., Stoyanov, E., Harari-Steinfeld, R., Steinfeld, I., Pappo, O., et al. (2014). Specific genomic and transcriptomic aberrations in tumors induced by partial hepatectomy of a chronically inflamed murine liver. *Oncotarget* 5, 10318–10331. doi: 10.18632/oncotarget.2515

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* 42, 1118–1125. doi: 10.1038/ng.717

Gella, A., Ponce, J., Cussó, R., and Durany, N. (2008). Effect of the nucleotides CMP and UMP on exhaustion in exercise rats. *J. Physiol. Biochem.* 64, 9–17. doi: 10.1007/BF03168230

Ghazalpour, A., Bennett, B. J., Shih, D., Che, N., Orozco, L., Pan, C., et al. (2014). Genetic regulation of mouse liver metabolite levels. *Mol. Syst. Biol.* 10:730. doi: 10.15252/msb.20135004

Gieger, C., Geistlinger, L., Altmaier, E., de Angelis, M. H., Kronenberg, F., Meitinger, T., et al. (2008). Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 4:e1000282. doi: 10.1371/journal.pgen.1000282

Hoefgen, R., and Nikiforova, V. J. (2008). Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiol. Plant* 132, 190–198. doi: 10.1111/j.1399-3054.2007.01012.x

Huang, X. Y., Huang, Z. L., Yang, J. H., Xu, Y. H., Sun, J. S., Zheng, Q., et al. (2016). Pancreatic cancer cell-derived IGFBP-3 contributes to muscle wasting. *J. Exp. Clin. Cancer Res.* 35:46. doi: 10.1186/s13046-016-0317-z

Hudson, N. J., Reverter, A., Greenwood, P. L., Guo, B., Cafe, L. M., and Dalrymple, B. P. (2015). Longitudinal muscle gene expression patterns associated with differential intramuscular fat in cattle. *Animal* 9, 650–659. doi: 10.1017/S1751731114002754

Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., et al. (2010). A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* 42, 137–141. doi: 10.1038/ng.507

Johnson, J. M., Yu, T., Strobel, F. H., and Jones, D. P. (2010). A practical approach to detect unique metabolic patterns for personalized medicine. *Analyst* 135, 2864–2870. doi: 10.1039/c0an00333f

Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., et al. (2012). Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* 28, 373–380. doi: 10.1093/bioinformatics/btr661

Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* 41, 56–65. doi: 10.1038/ng.291

Kim, T. H., Jun, H. Y., Kim, K. J., Lee, Y. H., Lee, M. S., Choi, K. H., et al. (2017). Hepatic alanine differentiates nonalcoholic steatohepatitis from simple steatosis in humans and mice: a proton MR spectroscopy study with long echo time. *J. Magn. Reson. Imag.* 46, 1298–1310. doi: 10.1002/jmri.25673

Kirchhof, P., Marijon, E., Fabritz, L., Li, N., Wang, W., Wang, T., et al. (2013). Overexpression of cAMP-response element modulator causes abnormal growth and development of the atrial myocardium resulting in a substrate for sustained atrial fibrillation in mice. *Int. J. Cardiol.* 166, 366–374. doi: 10.1016/j.ijcard.2011.10.057

Knight, J., Easter, L. H., Neiberg, R., Assimos, D. G., and Holmes, R. P. (2009). Increased protein intake on controlled oxalate diets does not increase urinary oxalate excretion. *Urol. Res.* 37, 63–68. doi: 10.1007/s00240-009-0170-z

Kowalik, M. A., Columbano, A., and Perra, A. (2017). Emerging role of the pentose phosphate pathway in hepatocellular carcinoma. *Front. Oncol.* 7:87. doi: 10.3389/fonc.2017.00087

Langfelder, P., and Horvath, S. (2008). WGCNA: an r package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Lee, H. S., Woo, S. J., Koh, H. W., Ka, S. O., Zhou, L., Jang, K. Y., et al. (2014). Regulation of apoptosis and inflammatory responses by insulin-like growth factor binding protein 3 in fibroblast-like synoviocytes and experimental animal models of rheumatoid arthritis. *Arthritis Rheumatol.* 66, 863–873. doi: 10.1002/art.38303

Maldonado, R., Smadja, C., Mazzucchelli, C., Sassone-Corsi, P., and Mazucchelli, C. (1999). Altered emotional and locomotor responses in mice deficient in the transcription factor CREM. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14094–14099. doi: 10.1073/pnas.96.24.14094

McMahon, G. M., Hwang, S. J., Clish, C. B., Tin, A., Yang, Q., Larson, M. G., et al. (2017). Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. *Kidney Int.* 91, 1426–1435. doi: 10.1016/j.kint.2017.01.007

Nicholson, G., Rantalainen, M., Li, J. V., Maher, A. D., Malmodin, D., Ahmadi, K. R., et al. (2011). A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.* 7:e1002270. doi: 10.1371/journal.pgen.1002270

Nordlie, R. C., Foster, J. D., and Lange, A. J. (1999). Regulation of glucose production by the liver. *Annu. Rev. Nutr.* 19, 379–406. doi: 10.1146/annurev.nutr.19.1.379

Ortega, F. J., Vazquez-Martin, A., Moreno-Navarrete, J. M., Bassols, J., Rodriguez-Hermosa, J., Gironés, J., et al. (2010). Thyroid hormone responsive Spot 14 increases during differentiation of human adipocytes and its expression is down-regulated in obese subjects. *Int. J. Obes.* 34, 487–499. doi: 10.1038/ijo.2009.263

Parks, B. W., Sallam, T., Mehrabian, M., Psychogios, N., Hui, S. T., Norheim, F., et al. (2015). Genetic architecture of insulin resistance in the mouse. *Cell Metab.* 21, 334–346. doi: 10.1016/j.cmet.2015.01.002

Passon, N., Puppin, C., Lavarone, E., Bregant, E., Franzoni, A., Hershman, J. M., et al. (2012). Cyclic AMP-response element modulator inhibits the promoter activity of the sodium iodide symporter gene in thyroid cancer cells. *Thyroid* 22, 487–493. doi: 10.1089/thy.2011.0360

Ponsuksili, S., Du, Y., Murani, E., Schwerin, M., and Wimmers, K. (2012). Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. *Genetics* 192, 1109–1122. doi: 10.1534/genetics.112.143081

Ponsuksili, S., Murani, E., Brand, B., Schwerin, M., and Wimmers, K. (2011). Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J. Lipid Res.* 52, 668–678. doi: 10.1194/jlr.M013342

Ponsuksili, S., Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., et al. (2015). Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. *PLoS One* 10:e0123678. doi: 10.1371/journal.pone.0123678

Ponsuksili, S., Trakooljul, N., Hadlich, F., Haack, F., Murani, E., and Wimmers, K. (2016). Genetically regulated hepatic transcripts and pathways orchestrate haematological, biochemical and body composition traits. *Sci. Rep.* 6:39614. doi: 10.1038/srep39614

Raup-Konsavage, W. M., Gao, T., Cooper, T. K., Morris, S. M. J., Reeves, W. B., and Awad, A. S. (2017). Arginase-2 mediates renal ischemia/reperfusion injury. *Am. J. Physiol. Renal. Physiol.* 313, F522–534. doi: 10.1152/ajprenal.00620.2016

Ryan, D., and Robards, K. (2006). Metabolomics: the greatest omics of them all? *Anal. Chem.* 78, 7954–7958. doi: 10.1021/ac0614341

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353–1358. doi: 10.1093/bioinformatics/bts163

Shin, S. Y., Fauman, E. B., Petersen, A. K., Krumsiek, J., Santos, R., Huang, J., et al. (2014). An atlas of genetic influences on human blood metabolites. *Nat. Genet.* 46, 543–550. doi: 10.1038/ng.2982

Sleiman, S. F., Henry, J., Al-Haddad, R., El Hayek, L., Abou Haidar, E., Stringer, T., et al. (2016). Exercise promotes the expression of brain derived neurotrophic factor (BDNF) through the action of the ketone body β-hydroxybutyrate. *eLife* 5:e15092. doi: 10.7554/eLife.15092

Staufner, C., Haack, T. B., Feyh, P., Gramer, G., Raga, D. E., Terrile, C., et al. (2016). Genetic cause and prevalence of hydroxyprolinemia. *J. Inherit. Metab. Dis.* 39, 625–632. doi: 10.1007/s10545-016-9940-2

Summitt, C. B., Johnson, L. C., Jönsson, T. J., Parsonage, D., Holmes, R. P., and Lowther, W. T. (2015). Proline dehydrogenase 2 (PRODH2) is a hydroxyproline dehydrogenase (HYPDH) and molecular target for treating primary hyperoxaluria. *Biochem. J.* 466, 273–281. doi: 10.1042/BJ20141159

van Kuilenburg, A. B., Dobritzsch, D., Meijer, J., Meinsma, R., Benoist, J. F., Assmann, B., et al. (2010). Dihydropyrimidinase deficiency: phenotype, genotype and structural consequences in 17 patients. *Biochim. Biophys. Acta* 1802, 639–648. doi: 10.1016/j.bbadis.2010.03.013

Vorobiev, S. M., Huang, Y. J., Seetharaman, J., Xiao, R., Acton, T. B., Montelione, G. T., et al. (2012). Human retinoblastoma binding protein 9, a serine hydrolase implicated in pancreatic cancers. *Protein Pept. Lett.* 19, 194–197. doi: 10.2174/092986612799080356

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17, 448–453. doi: 10.1038/nm.2307

Wang, Y. A., Sun, Y., Palmer, J., Solomides, C., Huang, L. C., Shyr, Y., et al. (2017). IGFBP3 modulates lung tumorigenesis and cell growth through IGF1 signaling. *Mol. Cancer Res.* 15, 896–904. doi: 10.1158/1541-7786.MCR-16-0390

Wu, H., Southam, A. D., Hines, A., and Vian, M. R. (2008). High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal. Biochem.* 372, 204–212. doi: 10.1016/j.ab.2007.10.002

Yabushita, S., Fukamachi, K., Tanaka, H., Fukuda, T., Sumida, K., Deguchi, Y., et al. (2013). Metabolomic and transcriptomic profiling of human K-ras oncogene transgenic rats with pancreatic ductal adenocarcinomas. *Carcinogenesis* 34, 1251–1259. doi: 10.1093/carcin/bgt053

Yang, S., Tschaplinski, T. J., Engle, N. L., Carroll, S. L., Martin, S. L., Davison, B. H., et al. (2009). Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genom.* 10:34. doi: 10.1186/1471-2164-10-34

Zhang, G., Saito, R., and Sharma, K. (2017). A metabolite-GWAS (mGWAS) approach to unveil chronic kidney disease progression. *Kidney Int.* 91, 1274–1276. doi: 10.1016/j.kint.2017.03.022

Zhou, L., Koh, H. W., Bae, U. J., and Park, B. H. (2015). Aggravation of post-ischemic liver injury by overexpression of insulin-like growth factor binding protein 3. *Sci. Rep.* 5:11231. doi: 10.1038/srep11231

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ponsuksili, Trakooljul, Hadlich, Methling, Lalk, Murani and Wimmers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 April 2019
doi: 10.3389/fgene.2019.00405

An Epigenome-Wide DNA Methylation Map of Testis in Pigs for Study of Complex Traits

*Xiao Wang and Haja N. Kadarmideen**

Quantitative Genomics, Bioinformatics and Computational Biology Group, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

Epigenetic changes are important for understanding complex trait variation and inheritance in pigs that are also a valuable biomedical model for human health research. Testis is the main organ for reproduction and boar taint in pigs; however, there have been no studies to-date on adult pig testis epigenome. The main objective of this study was to establish a genome-wide DNA methylation map of pig testis that would help identify candidate epigenetic biomarkers and methylated genes for complex traits such as male reproduction, fertility or boar taint. Reduced Representation Bisulfite Sequencing (RRBS) was used to study methylation levels of cytosine in nine pig testis samples. The results showed that genome-wide methylation status of nine samples overlapped greatly and their variation among pigs were low. The methylation levels of promoter, exon, intron, cytosine and guanine dinucleotide (CpG) islands and CpG island shores regions were 0.15, 0.47, 0.55, 0.39, and 0.53, respectively. Cytosines binding to CpG islands showed different methylation levels between exon and intron regions. All methylation levels of CpG islands were lower than CpG island shores in different genic features. The distribution of 12,738 differentially methylated cytosines (DMCs) within CpG islands, CpG island shores and other regions was 36.86, 21.65, and 41.49%, respectively, and was 0.33, 1.71, 5.95, and 92.01% in promoter, exon, intron and intergenic regions, respectively. Methylation levels of DMCs in promoter, exon and intron regions were significantly different between CpG islands and CpG island shores ($P < 0.05$). A total of 898 genes with 2089 DMCs were enriched in 112 Gene Ontology (GO) terms. Fifteen methylated genes from our study were associated with fertility or boar taint traits. Our analysis revealed the methylation patterns in different genic features and CpG island regions of testis in pigs, and summarized several candidate genes associated with DMCs and the involved GO terms. These findings are helpful to understand the relationship between DNA methylation and genic CpG islands, to provide candidate epigenetic regions or biomarkers for pig production and welfare and for translational epigenomic studies that use pigs as an animal model for human research.

Keywords: pig, testis, epigenome, DNA methylation, RRBS, DMC

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Zhe Zhang, South China Agricultural University, China
Ricardo Zanella, The University of Passo Fundo, Brazil

***Correspondence:**
Haja N. Kadarmideen
hajak@dtu.dk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *24 October 2018*
Accepted: *12 April 2019*
Published: *30 April 2019*

Citation:
Wang X and Kadarmideen HN (2019) An Epigenome-Wide DNA Methylation Map of Testis in Pigs for Study of Complex Traits. Front. Genet. 10:405. doi: 10.3389/fgene.2019.00405

Abbreviations: BGI, Beijing Genomics Institute; bp, base pair; cm, centimetre; CO2, Carbon dioxide; CpG, Cytosine and guanine dinucleotide; CTCF, CCCTC-binding factor; DMC, Differentially methylated cytosine; DMR, Differentially methylated region; FDR, False discovery rate; GO, Gene ontology; kb, kilo base pairs; kg, kilogram; Mb, mega base pair; mg, milligram; ml, millilitre; NGS, Next generation sequencing; PCR, Polymerase chain reaction; RNA-Seq, RNA sequencing; RRBS, Reduced representation bisulfite sequencing; SNP, Single nucleotide polymorphism; SSC, *Sus scrofa* chromosomes; TSS, Transcription start site; WGBS, Whole genome bisulfite sequencing.

INTRODUCTION

Pig is a valuable biomedical model of human obesity and metabolic diseases due to the anatomic, biochemical, pharmacological, pathological, and physiological similarities to the human (Kogelman et al., 2013; Kogelman and Kadarmideen, 2016). The previous study showed that the key role of epigenetic mechanisms in male gamete could widely affect human reproduction (Stuppia et al., 2015). Testis is the reproductive gland to produce sperm, so studying epigenetics of testis in pigs could improve our understanding of epigenetic molecular mechanisms related to male fertility and semen quality. Testis epigenome is also essential for the study of inheritance of boar taint in pigs – an unpleasant smell originating from cooking pork meat from uncastrated male pigs that is inherited (Strathe et al., 2013). Epigenetics is defined as changes in gene function that are heritable and no change in DNA sequence (Wu and Morris, 2001). As a major epigenetic modification, DNA methylation has been examined to be associated with growth (Jin et al., 2014), immune response (Wang et al., 2017), and reproduction traits (Bell et al., 2011) in pigs.

With high density of DNA methylation of cytosine and guanine dinucleotides (CpGs), CpG islands play an important role in gene regulation and transcriptional repression (Goldberg et al., 2007). The genome around the CpG islands can be widely affected by the methylation levels (Long et al., 2017). CpG island shores are strongly related to a specific tissue and are involved in modulating gene expression (Doi et al., 2009; Irizarry et al., 2009b). Most variable regions in terms of methylation such as methylation differences between tissues are CpG island shores rather than CpG islands themselves (Irizarry et al., 2009a; Hansen et al., 2011). DNA methylation in promoters is usually restricted to genes in a long-term stabilization of repressed states; therefore, promoter methylation can be a methylation inhibitor of therapeutic targets to silence genes (Yang et al., 2014). Most gene bodies are CpG-poor and extensively methylated, but their methylation can be a potential therapeutic target. Since DNA demethylation of the gene bodies could cause the down-regulation, so DNA methylation inhibitors can down regulate oncogenes and metabolic genes (Jones, 2012; Yang et al., 2014).

Reduced representation bisulfite sequencing (RRBS), based on next generation sequencing (NGS) technology, has been implemented to analyze patterns of DNA methylation by reducing the portion of the genome digestion (Meissner et al., 2005). Subsequently, reduced representation CpG sites are sequenced after restriction enzyme MspI digestion in CpG islands, promoters and enhancers (Smith et al., 2009). The RRBS method primarily focuses on the enrichment of CpG-rich regions rather than the non-CpG regions (Meissner et al., 2005). In mammals, DNA methylation almost exclusively occurs at CG dinucleotides with ratios of 70–80% throughout the genome (Ehrlich et al., 1982; Law and Jacobsen, 2010). Therefore, the information of CpG islands and gene-associated CpG sites can be provided by RRBS method (Choi et al., 2015). Currently, RRBS analysis of the pigs has been presented using intestinal tissue (Gao et al., 2014), ovaries (Yuan et al., 2016), and neocortex, liver, muscle and spleen (Choi et al., 2015).

Genome-wide DNA methylation patterns in porcine ovaries and porcine prepubertal testis have been profiled (Yuan et al., 2016; Chen et al., 2018), but to the best of our knowledge, genome-wide NGS-based methylation studies on adult testis epigenome in pigs have not been reported. The main objective of this study was to develop a map of DNA methylome for porcine testis using RRBS on nine testis samples of pigs and then characterize their methylome using bioinformatics methods. We characterized porcine adult testis epigenome by reporting the methylation levels and patterns in genic features and CpG islands for each testis sample. We identified differentially methylated cytosine (DMC) in nine sample to find DMC associated genes, and their involved Gene Ontology (GO) terms and pathways in pigs. Finally, we compared our results with other similar studies and provided a list of 15 candidate epigenetic biomarkers associated with male fertility (e.g., infertility, litter size, number of stillborn, and so on), boar taint (Skatole, Androstenone) and other complex traits linked to testis of pigs.

MATERIALS AND METHODS

Pig Samples

Nine commercial purebred Landrace male pigs with similar genetic background from nine different sire families were raised by the same *ad libitum* feeding of same feed type in the same farm/environment. All pigs were slaughtered at an age of around 22 weeks by carbon dioxide (CO_2) submersion at a commercial slaughterhouse (Danish Crown, Herning, Denmark), when they reached the slaughter weight of 105 kg. Testis tissue samples were retrieved by punch biopsy into the middle part of the testis with an inner punch distance of 2 cm. Thus, all of the testis samples were collected from the same part of the testis. Each sample weighed approximately 150 mg. These pigs were not treated by immunological castration or other castrating processes during the feeding period, so they had intact testis with normal fertility and viable sperms before or at slaughter.

Tissue samples were immediately immersed into the 1.5 ml RNAlater (QIAGEN, Hilden, Germany). All samples were stored at −20°C. Restriction enzyme digestion, adaptor ligation, size selection (40–220 bp fragments), bisulfite treatment, polymerase chain reaction (PCR) amplification and library construction were performed at BGI (Beijing Genomics Institute) Co., Ltd., Shenzhen, Guangdong, China. The nine samples were sequenced by a paired-end 100 bp flow cell in an Illumina HiSeq 2500 machine (PE-100bp FC; Illumina, San Diego, CA, United States) using RRBS method.

Quality Control, Read Alignment, and Trimming

RRBS adapters and reads less than 20 bases long were trimmed by Trimmomatic software (version 0.36) (Bolger et al., 2014). Then, Bismark Bisulfite Mapper (version 0.19.0) (Krueger and Andrews, 2011) was applied to map clean reads to the porcine reference genome (Sscrofa11.1/susScr11) downloaded from the UCSC

website[1], and the cytosine methylation status was determined accordingly. Bismark Bisulfite Mapper includes three steps: genome preparation, alignment using Bowtie 2 (version 2.3.3.1) (Langmead and Salzberg, 2012) and methylation extractor. Bismark methylation extractor outputs read coverage and methylation percentage of detected methylated or unmethylated reads at one genomic position. The numbers of methylated and unmethylated CpG and non-CpG (CHG and CHH, H representing A/C/T) sites were also calculated for each sample. The read coverages lower than 10 counts were trimmed for discarding the unqualified reads. If an experiment suffered from PCR duplication bias, some clonal reads will impair accurate determination of methylation. Thus, cytosines with a percentile of read coverage higher than the 99.9th were also discarded for each sample.

Genome-Wide DNA Methylation Levels and Methylation Patterns

The relationships of genome-wide methylation levels with densities of CpG islands, CpG island shores and genes were calculated through regression and correlation analysis, and counted by one mega base pairs (Mb) windows for each sample. Similarities and differences of genome structure, CpG islands and methylation level between genomic intervals were visualized by R package *RCircos* (version 1.2.0) (Zhang et al., 2013). Genic features were divided into promoter, exon and intron regions along the porcine genome. Afterward, we localized CpG islands and CpG island shores to these three genic features and investigated methylation patterns of genic CpG islands. Methylation patterns of CpG islands located at different genic features were visualized by R package *plot3D*.

Differentially Methylated Cytosine (DMC) and Annotation

Methylation levels of cytosines were analyzed by the R package *methylKit* (version 1.4.0) (Akalin et al., 2012) based on the Bismark coverage file. Genome-wide cytosine sites were combined into one object to obtain the locations covered in all nine samples. In this study, methylation level of nine samples were considered as nine treatment levels in the logistic regression model to calculate *P*-values, which were then adjusted to *Q*-values using false discovery rate (FDR) to account for multiple hypothesis testing (Storey and Tibshirani, 2003). Chi-squared (χ^2) test was used to determine the statistical significance of methylation differences between samples. Finally, we matched all DMCs into one file that included chromosomes, positions, *P*-values, *Q*-values, associated genes and their genic features, positions of CpG islands and CpG island shores and methylation levels of nine samples.

In this study, we defined CpG islands as a region with at least 200 bp, a GC fraction more than 0.5 and an observed-to-expected ratio of CpG more than 0.6. CpG island shores were then defined as regions of 2 kilo base pairs (kb) in length adjacent to CpG islands (Gardiner-Garden and Frommer, 1987). The CpG and DMC annotation within gene components of promoter, exon, intron and intergenic regions, and CpG islands, CpG island shores and other regions was performed using R package *genomation* (version 1.10.0) (Akalin et al., 2015). The porcine RefSeq and CpG island database (Sscrofa11.1/susScr11) for annotations were derived from the UCSC website[2].

Gene Ontology (GO) Enrichment and Pathway Analysis

GO enrichment and pathway analysis were analyzed in DAVID (Database for Annotation, Visualization and Integrated Discovery) Bioinformatics Resources 6.8[3]. NCBI reference sequences associated with DMCs were used in DAVID for the species of *Sus scrofa*. Significant GO terms and pathways were selected after filtering with $P < 0.01$. GO terms for the genes associated with DMCs were visualized by R package *GOplot* (version 1.0.2) (Walter et al., 2015).

RESULTS

Statistics of Alignment With Porcine Reference Genome

In this study, bisulfite conversion efficiencies of these nine samples ranged from 98 to 99%. The RRBS sequencing generated approximately 59,328,166 read pairs per sample. On average, 58,604,646 read pairs survived the pre-processing step. The 49% of the remaining read pairs was uniquely aligned to the porcine reference genome. The reads pairs were located in 9,006,052 sites, which meant that the average depth of RRBS sequencing reads and uniquely aligned reads were approximately equal to 13 and 6.5, respectively (**Table 1**). A total of 871,462,976 averaged cytosines were analyzed from 28,944,768 uniquely aligned reads pairs including methylated and unmethylated cytosines in CpG/CHG/CHH contexts (**Supplementary Table S1**). It revealed that a paired-end 100 bp read evenly contained 30 analyzed cytosines. Additionally, a per-sample CpG methylation rate ranged from 46 to 53%. The per-sample average percentages of cytosine methylation rate in CHG and CHH sites were 0.89 and 0.63%, respectively (**Table 1**).

It was obvious that the number of CpG sites was different at read coverage below 10, thus, the trimming criterion for read coverage was set at 10 (**Figure 1A**). **Figure 1B** revealed that the CpG site numbers of sample 1 and sample 9 were lower than the average value, while sample 5 has more CpG sites after trimming. Approximately, 9 million CpG sites were generated in each sample with read coverage equal to 21 (**Figure 1C**). After discarding coverage both lower than 10 and higher than 99.9th percentile, the averaged read coverage increased from 21 to 34, and the number of CpG sites reduced to a half (**Figure 1C**). The details of read coverages and methylation rates in CpG context of nine samples are listed in **Supplementary Table S2**. In addition, the coverage distributions per cytosine of nine samples after trimming are shown in **Supplementary Figure S1**. The

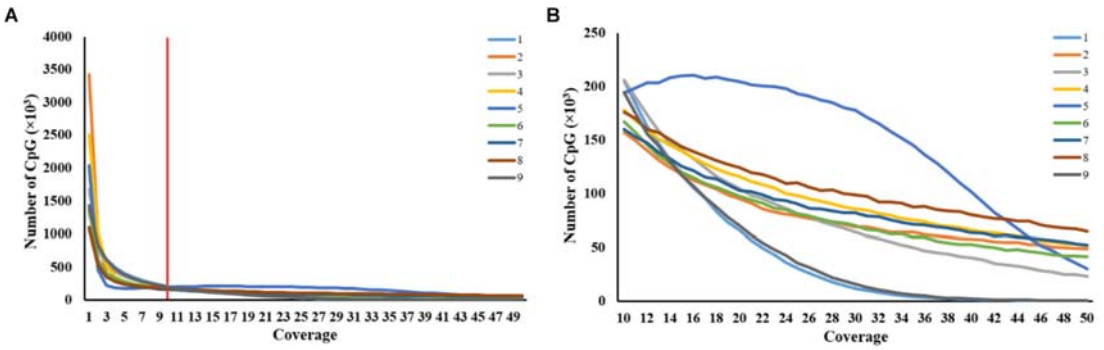
[1]http://hgdownload.cse.ucsc.edu/goldenPath/susScr11/bigZips/susScr11.fa.gz

[2]http://genome.ucsc.edu/cgi-bin/hgTables

[3]https://david.ncifcrf.gov/

TABLE 1 | Statistics of clean reads' alignment with porcine reference genome (Sscrofa11.1/susScr11) and methylation rates in CpG, CHG, and CHH contexts.

Sample	Clean read pair	Uniquely aligned rate	Number of aligned site	Total number of analyzed cytosine	Cytosine methylation rate in CpG context	Cytosine methylation rate in CHG context	Cytosine methylation rate in CHH context
1	16,505,578	46%	6,555,417	210,492,580	49%	0.91%	0.61%
2	93,817,089	51%	11,786,693	1,458,034,594	53%	0.99%	0.69%
3	38,026,074	47%	8,350,750	507,968,318	46%	0.84%	0.58%
4	75,769,839	51%	11,024,632	1,161,664,236	52%	0.87%	0.62%
5	57,267,890	51%	10,230,855	994,282,472	50%	0.68%	0.52%
6	68,607,455	46%	8,427,406	881,065,710	46%	0.89%	0.64%
7	85,068,927	49%	8,799,356	1,220,798,901	49%	0.92%	0.67%
8	75,438,276	51%	9,259,657	1,194,394,820	51%	0.92%	0.67%
9	16,940,690	47%	6,619,706	214,465,154	50%	0.95%	0.66%
Mean	58,604,646	49%	9,006,052	871,462,976	50%	0.89%	0.63%
SD	28,617,798	2.3%	1,798,552	456,951,421	2.4%	0.09%	0.05%



C

a | b	Coverage	Methylated	Unmethylated	Site in total	Trimmed coverage	Trimmed methylated	Trimmed unmethylated	Trimmed site in total
Mean (SD)	21 (9)	10 (5)	11 (4)	9,006,052 (1,798,552)	34 (12)	16 (6)	18 (6)	4,528,219 (1,776,393)

FIGURE 1 | Statistics of averaged coverage in CpG context. **(A)** Number of CpG sites at different coverage of original data. Note: Red line indicated the coverage at 10. **(B)** Number of CpG sites at different coverage of trimmed data. **(C)** Comparison of statistics of averaged coverage between original and trimmed data.

percent methylation distributions per cytosine of nine samples after trimming were shown through histograms on the diagonal of **Supplementary Figure S2**.

Genome-Wide DNA Methylation Status

The methylation levels against densities of CpG islands, CpG island shores and genes are shown in **Figure 2**. The genome-wide methylation status of nine samples showed the same trends and they overlapped greatly, suggesting that the biological variation between nine samples was low. Our analysis showed that the global CpG methylation rate was similar among the nine samples with Pearson's correlation scores ranging from 0.95 to 0.98 (**Supplementary Figure S2**). The methylation levels varied across the different chromosomes with higher methylation variation in regions of low gene abundance, whereas lower methylation variation in those of high gene abundance (**Figure 2**). The regression coefficients of densities of genes, CpG islands and CpG island shores on methylation level were −2.20 ($P < 0.001$), 59.04 ($P < 0.001$), and 73.65 ($P < 0.001$), respectively, on average, over nine samples (**Supplementary Figure S3** and **Supplementary Table S3**). The correlations between methylation levels and densities of genes, CpG islands and CpG island shores were -0.12, 0.25, and 0.23, respectively (**Supplementary Table S3**). These results suggested that genome hypomethylation in CpG islands was beneficial for the promotion of gene transcription, but their correlations were not so high.

Methylation Patterns of CpG Islands Located at Different Genic Features

To investigate the interaction of methylation levels between genes and CpG islands, we divided the porcine genome into three genic features (promoters, exons, and introns) and then localized CpG islands to these genic features. Methylation levels at different genic features and CpG islands displayed variously, with lowest values in the promoter regions. The methylation level were 0.15, 0.47, 0.55, 0.39, and 0.53 in the promoter, exon, intron, CpG islands, and CpG island shores regions, respectively, on average, over nine samples (**Figure 3A**). Comparisons of CpG islands and CpG island shores at different genic features revealed that the methylation levels of promoter regions were also the lowest.

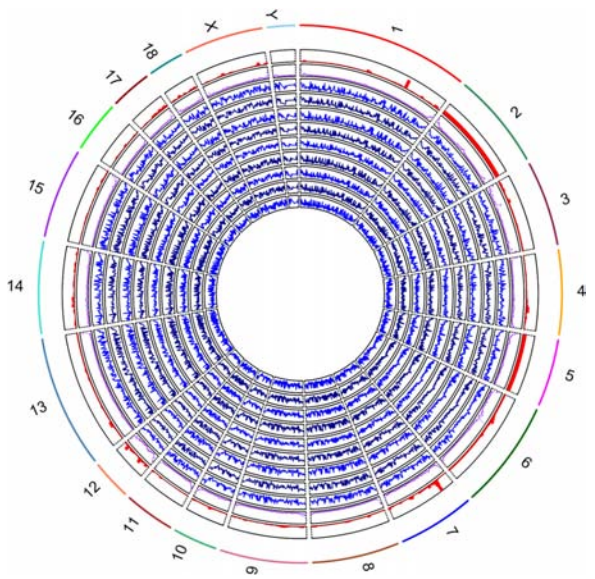


FIGURE 2 | Global methylation levels of nine samples was shown by lines of in blue (track 1, 2, 3, 4, 5, 6 7, 8, and 9) from inside to outside. The methylation levels and the densities of CpG islands by scatter plot in purple color (track 10), and genes by histograms in red color (track 11) were counted by 1 Mb windows. The labels of outside track represented the chromosomes of the porcine genome.

Meanwhile, CpG island shores located in intron regions showed slightly higher methylation levels than those located in exon regions, while CpG islands showed significant higher methylation levels (**Figures 3B,C**). Comparing with the methylation patterns in three different genic features, methylation levels of CpG islands were all lower than CpG island shores in the promoter, exon, and intron regions (**Figures 3D–F**).

Differentially Methylated Cytosines (DMC) and Annotations

A total of 1,244,043 CpG sites was covered in nine samples, and the number of identified DMCs was 12,738 with the level of $Q < 0.01$. Details of 12,738 DMCs with chromosomes, positions, *P*-values, *Q*-values, associated genes, genetic features and methylation levels are listed in **Supplementary File S1**. Percentages of 1,244,043 CpG sites annotated within promoter, exon, intron and intergenic regions were distributed as 5.33, 1.23, 3.80, and 89.64%, respectively. Additionally, the distribution of 1,244,043 CpG sites annotation within CpG islands, CpG island shores and other regions was 57.41, 14.71, and 27.88%, respectively. However, the distributions were 0.33, 1.71, 5.95, and 92.01% within promoter, exon, intron, and intergenic regions, respectively, when only considering the 12,738 DMCs. The distributions of DMCs annotated within CpG islands, CpG island shores and other regions were 36.86, 21.65, and 41.49%, respectively (**Figure 4**). The percentages of DMCs associated with CpG islands located in gene promoter, exon, intron, and intergenic regions were 69.05, 53.67, 32.32, and 36.72%, respectively. They were all higher than the DMCs associated with CpG island shores with the values of 19.05, 13.76, 24.01, and 21.66% in promoter, exon, intron, and intergenic regions, respectively (**Table 2**).

Among 19 ($n = 18 + 1$) *Sus scrofa* chromosomes (SSC), DMCs occupied SSC12 (12.1%) mostly, and nearly no DMCs occupied SSC X and SSC Y with the percentages of 0.4 and 0.1%, respectively (**Figure 5A**). DMCs were located mostly in the shorter genes and to lesser extent in the longer genes. Similarly, most of DMCs were located in CpG islands with a short length from 200 to 1000 bp (**Figure 5B**). Methylation levels of DMCs in different genic features were different, with the lowest values of CpG islands in the promoter regions. Student's *t*-tests showed that methylation levels of DMCs in promoter, exon and intron regions were significantly different between CpG islands and CpG island shores ($P < 0.05$), while those of intergenic regions were extremely significant ($P < 0.001$) (**Figure 5C**). The averaged methylation levels on different chromosomes and different individuals were similar, with values close to 50% (**Figure 5D**).

Genes Associated With DMCs and Their Gene Ontology (GO) Enrichment and Pathway Analyses

We found that 976 DMCs were annotated within gene components of 415 genes after matching 12,738 DMCs to the porcine RefSeq database (Sscrofa11.1/susScr11) (**Supplementary File S1**). Fifteen genes associated with DMCs found to be related to fertility or boar taint traits were also reported by other studies (**Table 3**). Genes *ACACA*, *CYP21A2*, *CYP27A1*, *HSD17B2*, *LHB*, *PARVG,* and *SERPINC1* were associated with boar taint, while genes *DICER1*, *PCK1*, *SS18,* and *TGFB3* were associated with pig reproduction traits. In addition, the other five genes (*CAPN10*, *FTO*, *HSD17B2*, *IGF2*, and *SALL4*) were found to be associated with fertility traits in human, in which *HSD17B2* also played a role in boar taint (**Table 3**).

Hereafter, 898 genes (296 unique genes) associated with 2089 DMCs (704 unique DMCs) were enriched in 112 GO terms (**Supplementary File S2**). The significant GO terms ($P < 0.01$) are shown with the texts including 7 GO terms of biological process, 5 GO terms of cellular component and 7 GO terms of molecular function (**Figure 6**). Generally, as more genes were enriched in the GO terms, the number of included DMCs increased (**Figure 6**). Two GO terms (GO: 0005737 and GO: 0005634) in the cellular component contained the genes and DMCs mostly, that were 80 and 78 enriched genes associated with 185 and 182 DMCs, respectively (**Supplementary File S2**). The 23 significant pathways ($P < 0.01$) are listed in **Supplementary Table S4**. The most significant pathway was insulin signaling pathway ($P = 9.89 \times 10^{-7}$) containing 16 genes namely *PHKG2*, *FASN*, *PHKG1*, *ACACA*, *IKBKB*, *FBP1*, *GYS1*, *PRKCZ*, *PRKAA2*, *PRKAG1*, *PCK1*, *ACACB*, *PIK3R5*, *SREBF1*, *AKT2,* and *MAP2K1* (**Supplementary Table S4**).

DISCUSSION

Generally, the bisulfite conversion rates ranged from 90 to 100%, but some conversion rates varied between 99 and 100%

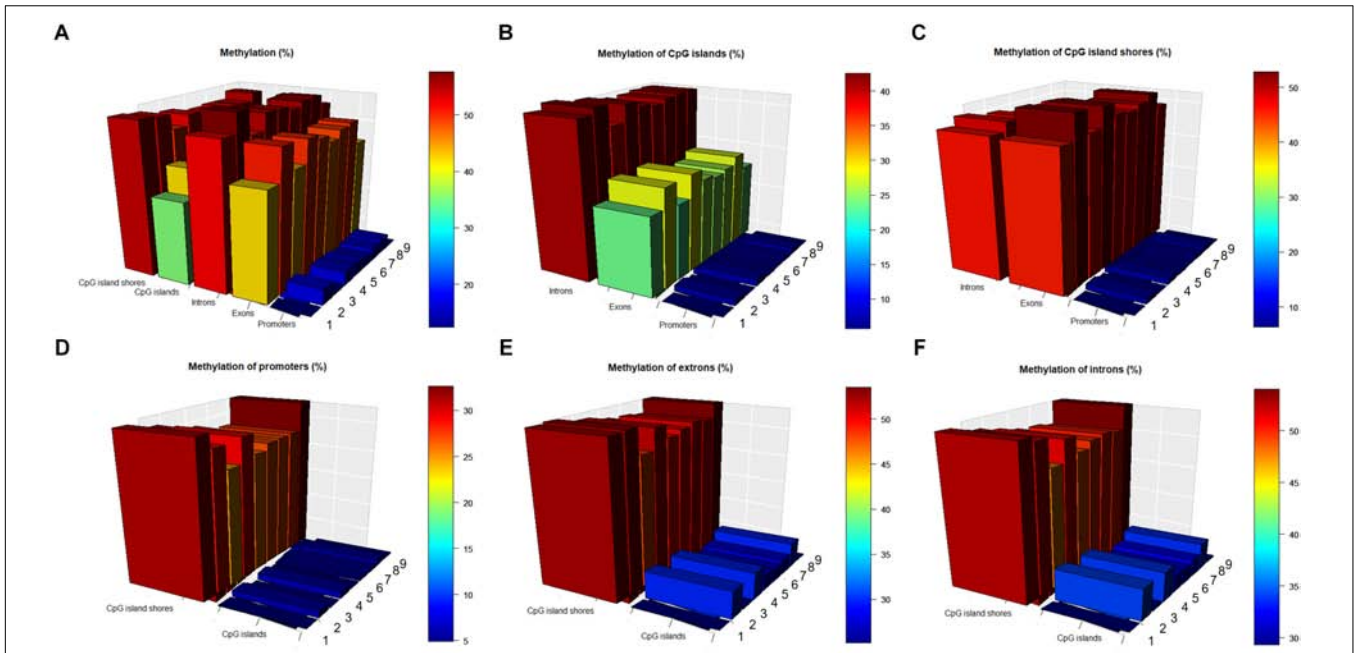


FIGURE 3 | Methylation patterns in different genic features and CpG islands regions. **(A)** Methylation levels (in %) at different genic features, CpG islands and CpG island shores. **(B)** Methylation levels (in %) of CpG islands at different genic features. **(C)** Methylation levels (in %) of CpG island shores at different genic features. **(D)** Methylation levels (in %) of promoters in the CpG islands and CpG island shores. **(E)** Methylation levels (in %) of exons in the CpG islands and CpG island shores. **(F)** Methylation levels (in %) of introns in the CpG islands and CpG island shores.

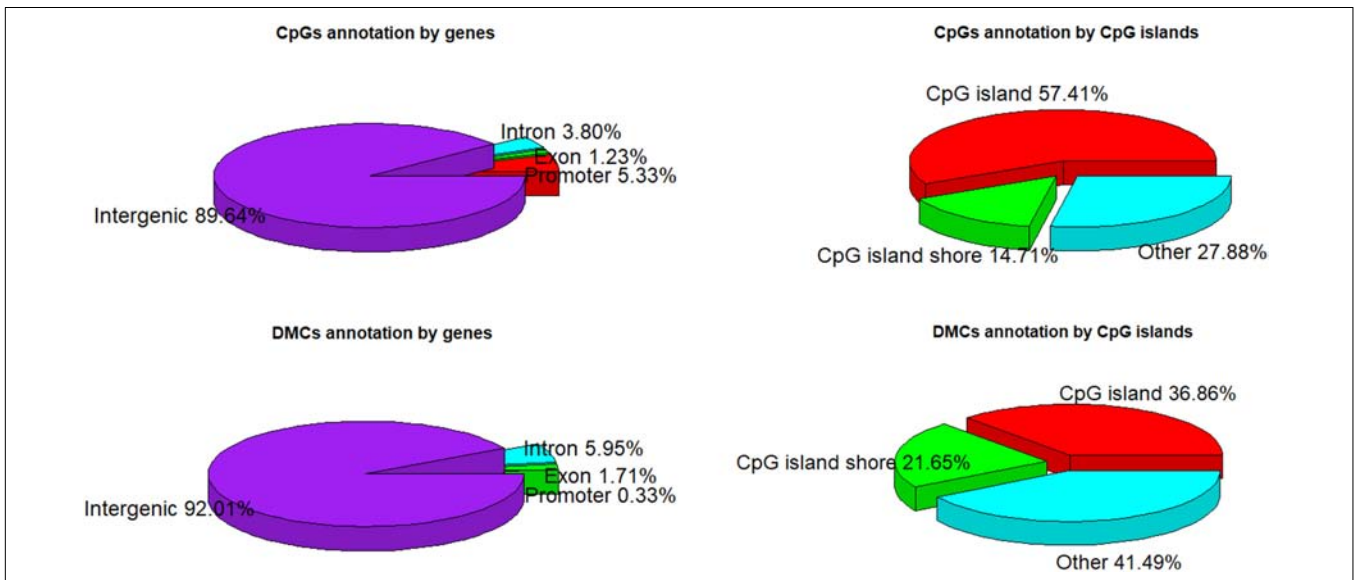


FIGURE 4 | CpGs and DMCs annotation by genes and CpG islands.

TABLE 2 | DMCs associated with CpG island regions located at different genic features.

Genic feature	CpG island				CpG island shore			
	Promoter	Exon	Intron	Intergenic	Promoter	Exon	Intron	Intergenic
Number	29	117	245	4304	8	30	182	2538
Percentage	69.05%	53.67%	32.32%	36.72%	19.05%	13.76%	24.01%	21.66%

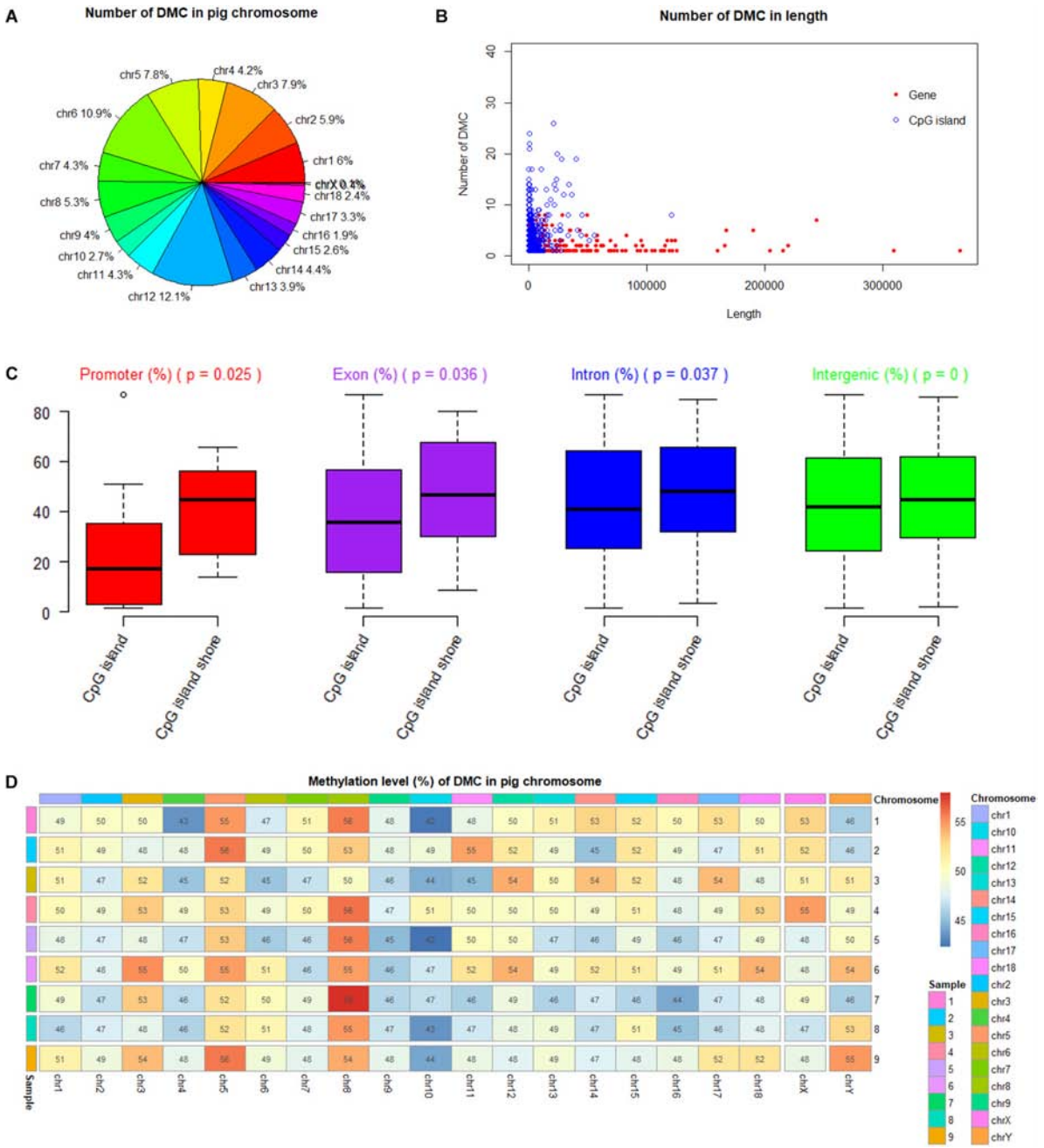


FIGURE 5 | Methylation status of DMC in porcine chromosomes, genic features and CpG islands. **(A)** Number of DMC in different pig chromosome. **(B)** Number of DMC in the different lengths of genes and CpG islands. **(C)** Comparison of methylation levels between CpG islands and CpG island shores at different genic feature with Student's *t*-tests. **(D)** Methylation levels of DMC in different pig chromosome.

depending on the commercial methods (Worm Ørntoft et al., 2017). This study showed higher bisulfite conversion efficiencies between 98 and 99%. A mapping efficiency of 38.3% was previously reported in RRBS sequencing of lamb muscle with fragment sizes of 50–150 bp, which increased to 61.4% with fragment sizes of 150–250 bp (Doherty and Couldrey, 2014). Similarly, our study revealed efficiency of 49% using 40–220 bp sizes that were uniquely mapped to the porcine reference genome (**Table 1**). It is consistent with 60% mapping rates using 110–220 bp sizes in RRBS sequencing for porcine ovaries (Yuan et al., 2016). We found that global CpG methylation levels ranged from 45 to 53% (50% on average), which is similar with other studies on pig methylation research using RRBS method (Gao et al., 2014; Choi et al., 2015; Schachtschneider et al., 2015), whereas non-CpG methylation levels (CHG and CHH sites) were less than 1% (**Table 1**). This is reasonable because

CpGs within poor-CpG regions are scarcely covered based on restriction enzyme digestion by the RRBS method (Meissner et al., 2005). Our results also showed 72% of CpG methylations were mapped to CpG islands (57.41%) and to CpG island shores (14.71%), that were higher than those of Choi's study (Choi et al., 2015). Whole genome bisulfite sequencing (WGBS) technology can produce many reads in poorly assembled non-coding DNA regions, resulting in lower mapping efficiency than RRBS method (Doherty and Couldrey, 2014). However, RRBS data sets have a somewhat lower average methylation level than WGBS data sets, because large stretches of repeat regions in non-coding DNA regions are generally highly methylated (Bird, 2002). Practically, some CpG sites had low coverage (1∼ 10) or are not even sequenced by the WGBS method, although all sites should be theoretically covered (Sun et al., 2015). Thus, average read depths of RRBS sequencing were higher than 10 in this study (**Table 1** and **Supplementary Table S1**) and in other studies (Zhao et al., 2016; Carmona et al., 2017). Overall, RRBS method remained a better choice when considering sequencing cost, read coverage and sufficient methylation information (Choi et al., 2015).

In many cell types of different species, percentages of methylations would have a bimodal distribution, which denoted that the majority of bases has either high or low methylation to indicate a site specificity (Ehrlich et al., 1982). This bimodal pattern was a possible function to keep the factor-mediated basal transcription profile of the preimplantation embryo (Cedar and Bergman, 2012). The CpG methylation percentage distribution would be measured with two peaks at 0 and 100%, when a large number of the CpG sites were sequenced in either unmethylated or fully methylated status (Falckenhayn et al., 2013; Zhang et al., 2017). Bimodal distribution is also an important metric to help reveal whether the experiments suffer from PCR duplication bias. If there is a high degree of clonal reads from PCR, some reads will be asymmetrically amplified and read coverage distribution will have a secondary peak correspondingly on the right side. This situation will impair accurate determination of percent methylation scores for those regions. Hence, this study discarded cytosines with a percentile of read coverage higher than 99.9th, and then showed the reasonably bimodal distribution (**Supplementary Figure 2**) in consistency with other results using different tissues in pigs (Choi et al., 2015).

Not only did DNA methylation have a correlation with gene transcription, but also the presence of methyl moieties inhibited gene expression *in vivo* (Razin and Cedar, 1991). It was suggested by our study that the regression coefficients and correlation coefficients of genes and methylation levels were both negative, ranging from −1.97 to −2.46 and from −0.10 to −0.14, respectively (**Supplementary Table S3**). In practice, the correlation coefficient between gene expression and methylation level was approximately 0.3, negative (Bock, 2012). Methylated genes might be associated with genomic region-specific DNA methylation patterns (Raza et al., 2017), and therefore, this study investigated promoter, exon and intron regions along the porcine genome and localized CpG islands to these genic features. The interactions of methylations between three genic features and CpG islands suggested that methylation levels of promoter regions were lowest in both CpG islands and CpG island shores (**Figure 3A**). It was well known that DNA methylation in a promoter was correlated with the transcription of a target gene (Niesen et al., 2005). Methylation levels of CpG islands were lower than CpG island shores in the promoter, exon and intron regions in this study (**Figures 3D–F**). These results demonstrated that CpG islands located in different genic features displayed effects on the methylation patterns of the associated genes. Irizarry et al. (Irizarry et al., 2009b) revealed a strong relation between methylations in CpG island shores located within 2 kb of an annotated transcription start site (TSS) and expression of associated genes. Meanwhile, CpG islands located in exon regions showed different methylation level with those located in intron regions (**Figures 3B,C**), which suggested that exons had an effect on the methylation patterns of CpG islands. Chen et al. (2018) has profiled methylation patterns for porcine testis at three prepubertal age points (i.e., 1, 2, and 3 months). They found that the methylation levels of promoters and CpG islands decreased as the pig gradually matured, while methylation levels of gene body kept stable (Chen et al., 2018). It was suggested that lower methylations in promoters could be a specific pattern for testis tissue in adult pig, because spermatogenic cells tended to be activated for the increasing gene expression requirement at this stage. Additionally, Yuan et al. (2017) revealed that CpG islands show lower methylation levels compared to their CpG island shore regions in porcine hypothalamus-pituitary-ovary axis. Methylation levels in introns, exons, and promoters gradually decreased both in CpG islands and CpG island shores (Yuan et al., 2017). The methylation patterns of hypothalamus-pituitary-ovary axis were similar to our results except that exons located in CpG island shores of this study showed slightly higher methylations than those located in CpG islands (**Figures 3B,C**).

The percentages of DMCs annotation within exon, intron and intergenic regions increased, whereas DMCs annotation within promoter region decreased dramatically, when comparing DMCs with CpGs annotation within genic features. Similarly, the percentage of DMCs annotation within CpG island shores increased, while DMCs annotation within CpG islands decreased (**Figure 4**). As Maunakea et al. (2010) found that the methylated CpG islands in 5′ promoter regions were less than 3%, DMCs found in promoter regions were also less than 1% in this study (**Figure 4**). The most common promoter type in the vertebrate genome was annotated gene promoters with the CpG islands and they occupied at above 70% (Saxonov et al., 2006). We found that approximately 69% of DMCs associated with CpG islands were located in promoter regions (**Table 2**). Liu et al. (Liu et al., 2017) reported that the proportions of hypermethylated CpG sites located in CpG islands, CpG shores and other locations were 25.49∼34.23%, 21.57∼40.75%, and 25.02∼52.94%, respectively, during different stages of human embryonic stem cells. Genes that contained differentially methylated regions (DMRs) in their first intron were more than the genes that contained DMRs in their promoter and their first exon (Anastasiadi et al., 2018), which are the same trend as this study (**Supplementary File S1**).

In humans, more than 80% of sperm cells were mainly composed in the testis (Bellve et al., 1977). The epigenetic modifications of germ cells occurring in the meiotic and post-meiotic phases of spermatogenesis are crucial for embryonic

TABLE 3 | Comparisons between DMC related genes of this study and identified genes for fertility or boat taint traits of other studies.

	This study				Other study				
Co-found gene	**Chromosome**	**Number of DMC**	**DMC position**	**Feature region**	**Specie**	**Tissue**	**Data type**	**Associated trait**	**References**
ACACA	SSC12	1	38585187	Intron	Pig	Testis	RNA-Seq	Androstenone	Moe et al., 2007
CAPN10	SSC15	1	139564776	Exon	Human	Leukocytes	SNP	Polycystic ovarian syndrome	Gonzalez et al., 2003
CYP21A2	SSC7	1	24086506	Intron and CpG island shore	Pig	Blood, leukocytes or semen	SNP array	Androstenone, testosterone, 17β-estradiol and estronsulphate.	Grindflek et al., 2011
CYP27A1	SSC15	1	120810825	Intron and CpG island	Pig	Testis	RNA-Seq	Boar taint	Drag et al., 2017
DICER1	SSC7	2	116375790, 116375791	Intron and CpG island shore	Pig	Testis and oviduct	RNA-Seq	Reproduction	Fischer et al., 2015
FTO	SSC6	2	31317922, 31317957	Intron	Human	Blood	SNP array	Polycystic ovarian syndrome	Barber et al., 2008
HSD17B2	SSC6	1	6296641	Exon	Pig/Human	Liver/ Endometrium	RNA-Seq /RNA	Skatole metabolism /Endometriosis	Zeitoun et al., 1998; Gunawan et al., 2013
IGF2	SSC2	1	1486537	Intron and CpG island shore	Human	Semen	DNA methylation	Male infertility	Poplinski et al., 2010
LHB	SSC6	1	54264270	Exon and CpG island shore	Pig		SNP array	Androstenone metabolism	Duijvesteijn et al., 2010
PARVG	SSC5	4	4874783, 4874934, 4874946, 4886005	Intron	Pig	Testis	RNA-Seq	Androstenone	Moe et al., 2007
PCK1	SSC17	1	57933961	Exon and CpG island shore	Pig	Placenta	RNA-Seq	litter size	Kwon et al., 2016
SALL4	SSC17	3	53089645, 53089693, 53090863	Exon and CpG island	Human	Peripheral blood leukocytes	SNP	Premature Ovarian Failure	Wang et al., 2009
SERPINC1	SSC9	1	116186998	Exon	Pig	Liver	RNA-Seq	Boar taint	Drag et al., 2017
SS18	SSC6	1	110725565	Intron	Pig	Ear	Whole genome	Number of stillborn	Onteru et al., 2012
TGFB3	SSC7	1	99138369	Exon	Pig		SNP array	Number of teats	Verardo et al., 2016

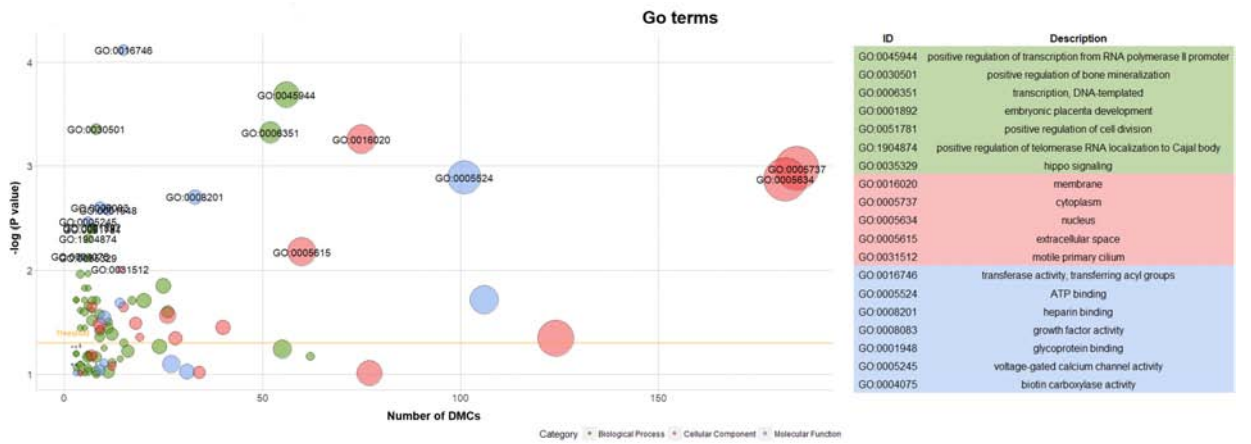


FIGURE 6 | Go term analysis of genes associated with DMCs. Note: Yellow line in the left panel meant the threshold of significant GO terms ($P < 0.05$).

development after fertilization (Marques et al., 2010). Due to the failure of re-methylation in spermatogonia or alterations to methylation maintenance in spermatocytes, sperm cells or the mature sperm cells, the abnormal DNA methylation patterns were observed in the infertile men (Cui et al., 2016). Therefore, the methylation patterns in genic features and CpG islands of pig testis were investigated to reveal significant cytosines and associated genes for epigenetic molecular mechanisms related to male fertility. Langenstroth-Röwer et al. (2017) used the marmoset monkey as the human model for testicular methylation study. They found that cytosines were predominantly unmethylated at regulatory regions of *H19*, *LIT1*, *SNRPN*, *MEST,* and *OCT4* in the germ cells. Meanwhile, DNA methylation pattern of *H19*, *MEST*, *DDX-4,* and *MAGE-A4* did not change in germ cell fractions (Langenstroth-Röwer et al., 2017). The genome-wide promoter methylation profiles identified 367 testis and epididymis-specific hypomethylated genes and 134 hypermethylated genes, many of them were involved in the GO terms of male reproduction (Wu et al., 2013). Compared with the fertile males, it was reported that a low methylation or unmethylation pattern at the *H19* was associated with hypermethylation at the *MEST* and a reduced sperm quality in the oligospermic patients (Niemitz and Feinberg, 2004). DMRs located in the upstream of TSS of the *H19* harbored several CCCTC-binding factor (CTCF) binding sites (Takai, 2001). However, CTCF binding to the maternal unmethylated DMR could prevent *IGF2* from accessing the common enhancers, and thus silencing its expression (Marques et al., 2010). Rajender et al. (2011) summarized that genes *MTHFR, PAX8, NTF3, SFN, HRAS, JHM2DA, IGF2, H19, RASGRF1, GTL2, PLAG1, D1RAS3, MEST, KCNQ1, LIT1,* and *SNRPN* were associated with male infertility. Our study also identified the DMCs located in the intron regions of *IGF2* (**Table 3**), which was involved in GO terms of positive regulation of cell division (GO: 0051781), extracellular space (GO: 0005615), and growth factor activity (GO: 0008083) (**Supplementary File S2**).

Our study revealed the methylation patterns in different genic features such as promotor, exon, intron and intergenic regions, as well as CpG islands, CpG island shores regions. Furthermore, our study reported many candidate genes harboring DMCs and the involved GO terms of testis in pig. Until now, several studies have concluded the important genes associated with male fertilities using SNP array, RNA-Seq datasets for humans (**Table 3**), however, epigenetic studies in pigs relating to male fertility are rare. This study has reported for the first time, DNA methylome (epigenomic) architecture in adult pig testis for study of male fertility in pigs. These results will also be useful for the study of boar taint in pigs associated with sensory meat quality, as boar taint is inherited and shows complex gene regulation patterns (Strathe et al., 2013; Drag et al., 2018). Since this study is based on sequence-level resolution of transmittable epigenetic changes, we believe it may also contribute to understanding and capturing part of the genetic variation that are not captured by SNP arrays (considered missing or "missing heritability") in genome-wide genomic prediction studies. As pig is a valuable biomedical model of human, the findings of this study are also very helpful to understand the relationship between DNA methylation and genic CpG islands, and provide candidate epigenetic biomarkers for the translational studies in human research.

CONCLUSION

This is the first study to report catalog of adult pig testis epigenome by developing a genome-wide DNA methylation map with the use of RRBS technology. We found that the methylation rates were lowest in promoters (0.15) and highest in introns (0.55). Cytosines binding to CpG islands showed different methylation patterns between intron and exon regions. Methylation levels of CpG islands were lower than CpG island shores in different genic features. We detected 12,738 DMCs in total. They distributions of DMCs within CpG islands, CpG island shores and other regions were 36.86, 21.65, and 41.49%, respectively. The distributions of DMCs were 0.33, 1.71, 5.95, and 92.01% in promoter, exon, intron and intergenic regions, respectively. Fifteen genes with DMCs were associated

with human fertility (*ACACA*, *CYP21A2*, *CYP27A1*, *HSD17B2*, *LHB*, *PARVG,* and *SERPINC1*), pig reproduction (*DICER1*, *PCK1*, *SS18,* and *TGFB3*) and boar taint traits (*CAPN10*, *FTO*, *HSD17B2*, *IGF2*, and *SALL4*). These findings on genome-wide epigenetic signatures will be useful to understand testis-related trait inheritance in pigs (e.g., male fertility, semen quality, boar taint) for pig production and welfare. This study, based on sequence-level resolution of epigenetic changes, also contributes to understanding and capturing part of the genetic variation that are considered missing ("missing heritability") in genome-wide genomic prediction studies. Since pigs are useful as an animal model for human research, epigenetic architecture of pigs would help in translational research.

ETHICS STATEMENT

Animal Care and Use Committee approval was not obtained for this study, because tissue samples were obtained from a commercial slaughter facility.

AUTHOR CONTRIBUTIONS

HK conceived, designed and implemented the epigenomic experiments including collection of tissue samples and processing of samples for methylome sequencing by RRBS, and improved the manuscript. XW analyzed the data. XW and HK interpreted the results and wrote the manuscript. Both authors read and approved the final manuscript.

FUNDING

Ph.D. Project funded by the Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark.

ACKNOWLEDGMENTS

We thank staff at the Slaughter house – Danish Crown in Herning, Denmark and Dr. Ruta Skinkyté-Juskiené for assistance in tissue collection and processing of samples. We thank Dr. Markus Drag for organizing shipments of samples for sequencing. XW received Ph.D. stipends from the Technical University of Denmark, DTU Bioinformatics and DTU Compute, Denmark, and the China Scholarship Council, China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00405/full#supplementary-material

FIGURE S1 | Histograms of log10 of read coverage per CpG site.

FIGURE S2 | Correlation analysis of the global CpG methylation patterns among nine samples. Note: Colors in the scatter plot indicate the number of CpG sites with identical methylation pattern (methylated or non-methylated): yellow denotes many correlations, blue denotes lack of correlation and green denotes different methylation patterns. Numbers in the upper right side represent the pairwise Pearson's correlation scores. Histograms on the diagonal are methylation distribution per CpG site for each sample.

FIGURE S3 | Regression of densities of genes, CpG islands and CpG island shores on methylation levels from one sample, all counted by 1 Mb windows.

TABLE S1 | Total number of aligned cytosine methylation in different contexts.

TABLE S2 | Statistics of coverage and methylation rates in CpG context.

TABLE S3 | Regression and correlation analysis of densities of genes, CpG islands and CpG island shores on methylation levels, all counted by 1 Mb windows.

TABLE S4 | Significant pathways ($P < 0.01$).

FILE S1 | Details of 12,738 DMCs with chromosomes, positions, *P*-values, *Q*-values, associated genes, genetic features and methylation levels.

FILE S2 | GO terms and enriched genes associated with DMCs.

REFERENCES

Akalin, A., Franke, V., Vlahoviček, K., Mason, C. E., and Schübeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* 31, 1127–1129. doi: 10.1093/bioinformatics/btu775

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13:R87. doi: 10.1186/gb-2012-13-10-R87

Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11:37. doi: 10.1186/s13072-018-0205-1

Barber, T. M., Bennett, A. J., Groves, C. J., Sovio, U., Ruokonen, A., Martikainen, H., et al. (2008). Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia* 51, 1153–1158. doi: 10.1007/s00125-008-1028-6

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., et al. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 12:R10. doi: 10.1186/gb-2011-12-1-r10

Bellve, A. R., Cavicchia, J. C., Millette, C. F., O'Brien, D. A., Bhatnagar, Y. M., and Dym, M. (1977). Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J. Cell Biol.* 74, 68–85. doi: 10.1083/jcb.74.1.68

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13, 705–719. doi: 10.1038/nrg3273

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Carmona, J. J., Accomando, W. P., Binder, A. M., Hutchinson, J. N., Pantano, L., Izzi, B., et al. (2017). Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans. *npj Genomic Med.* 2:13. doi: 10.1038/s41525-017-0012-9

Cedar, H., and Bergman, Y. (2012). Programming of DNA methylation patterns. *Annu. Rev. Biochem.* 81, 97–117. doi: 10.1146/annurev-biochem-052610-091920

Chen, X., Shen, L. H., Gui, L. X., Yang, F., Li, J., Cao, S. Z., et al. (2018). Genome-wide DNA methylation profile of prepubertal porcine testis. *Reprod. Fertil. Dev.* 30, 349–358. doi: 10.1071/RD17067

Choi, M., Lee, J., Le, M. T., Nguyen, D. T., Park, S., Soundrarajan, N., et al. (2015). Genome-wide analysis of DNA methylation in pigs using reduced representation bisulfite sequencing. *DNA Res.* 22, 343–355. doi: 10.1093/dnares/dsv017

Cui, X., Jing, X., Wu, X., Yan, M., Li, Q., Shen, Y., et al. (2016). DNA methylation in spermatogenesis and male infertility. *Exp. Ther. Med.* 12, 1973–1979. doi: 10.3892/etm.2016.3569

Doherty, R., and Couldrey, C. (2014). Exploring genosme wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Front. Genet.* 5:126. doi: 10.3389/fgene.2014.00126

Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41, 1350–1353. doi: 10.1038/ng.471

Drag, M., Hansen, M. B., and Kadarmideen, H. N. (2018). Systems genomics study reveals expression quantitative trait loci, regulator genes and pathways associated with boar taint in pigs. *PLoS One* 13:e0192673. doi: 10.1371/journal.pone.0192673

Drag, M., Skinkyté-Juskiené, R., Do, D. N., Kogelman, L. J. A., and Kadarmideen, H. N. (2017). Differential expression and co-expression gene networks reveal candidate biomarkers of boar taint in non-castrated pigs. *Sci. Rep.* 7:12205. doi: 10.1038/s41598-017-11928-0

Duijvesteijn, N., Knol, E. F., Merks, J. W. M., Crooijmans, R. P. M. A., Groenen, M. A. M., Bovenhuis, H., et al. (2010). A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genet.* 11:42. doi: 10.1186/1471-2156-11-42

Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., Mccune, R. A., et al. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.* 10, 2709–2721. doi: 10.1093/nar/10.8.2709

Falckenhayn, C., Boerjan, B., Raddatz, G., Frohme, M., Schoofs, L., and Lyko, F. (2013). Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. *J. Exp. Biol.* 216(Pt 8), 1423–1429. doi: 10.1242/jeb.080754

Fischer, D., Laiho, A., Gyenesei, A., and Sironen, A. (2015). Identification of reproduction-related gene polymorphisms using whole transcriptome sequencing in the large white pig population. *G3* 5, 1351–1360. doi: 10.1534/g3.115.018382

Gao, F., Zhang, J., Jiang, P., Gong, D., Wang, J. W., Xia, Y., et al. (2014). Marked methylation changes in intestinal genes during the perinatal period of preterm neonates. *BMC Genomics* 15:716. doi: 10.1186/1471-2164-15-716

Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282. doi: 10.1016/0022-2836(87)90689-9

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635–638. doi: 10.1016/j.cell.2007.02.006

Gonzalez, A., Abril, E., Roca, A., Aragón, M. J., Figueroa, M. J., Velarde, P., et al. (2003). Specific CAPN10 gene haplotypes influence the clinical profile of polycystic ovary patients. *J. Clin. Endocrinol. Metab.* 88, 5529–5536. doi: 10.1210/jc.2003-030322

Grindflek, E., Lien, S., Hamland, H., Hansen, M. H. S., Kent, M., van Son, M., et al. (2011). Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. *BMC Genomics* 12:362. doi: 10.1186/1471-2164-12-362

Gunawan, A., Sahadevan, S., Cinar, M. U., Neuhoff, C., Große-Brinkhaus, C., Frieden, L., et al. (2013). Identification of the novel candidate genes and variants in boar liver tissues with divergent skatole levels using RNA deep sequencing. *PLoS One* 8:e72298. doi: 10.1371/journal.pone.0072298

Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43, 768–775. doi: 10.1038/ng.865

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., et al. (2009a). Genome-wide methylation analysis of human colon cancer reveals similar hypo-and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41:178. doi: 10.1038/ng.298.Genome-wide

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., et al. (2009b). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186. doi: 10.1038/ng.298

Jin, L., Jiang, Z., Xia, Y., Lou, P., Chen, L., Wang, H., et al. (2014). Genome-wide DNA methylation changes in skeletal muscle between young and middle-aged pigs. *BMC Genomics* 15:653. doi: 10.1186/1471-2164-15-653

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230

Kogelman, L. J. A., and Kadarmideen, H. N. (2016). Applications of systems genetics and biology for obesity using pig models. *Syst. Biol. Anim. Prod. Heal.* 1, 25–42. doi: 10.1007/978-3-319-43335-6_2

Kogelman, L. J. A., Kadarmideen, H. N., Mark, T., Karlskov-Mortensen, P., Bruun, C. S., Cirera, S., et al. (2013). An F2 pig resource population as a model for genetic studies of obesity and obesity-related diseases in humans: design and genetic parameters. *Front. Genet.* 4:29. doi: 10.3389/fgene.2013.00029

Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167

Kwon, S. G., Hwang, J. H., Park, D. H., Kim, T. W., Kang, D. G., Kang, K. H., et al. (2016). Identification of differentially expressed genes associated with litter size in Berkshire pig placenta. *PLoS One* 11:e0153311. doi: 10.1371/journal.pone.0153311

Langenstroth-Röwer, D., Gromoll, J., Wistuba, J., Tröndle, I., Laurentino, S., Schlatt, S., et al. (2017). De novo methylation in male germ cells of the common marmoset monkey occurs during postnatal development and is maintained in vitro. *Epigenetics* 12, 527–539. doi: 10.1080/15592294.2016.1248007

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719

Liu, Z., Feng, Q., Sun, P., Lu, Y., Yang, M., Zhang, X., et al. (2017). Genome-wide DNA methylation drives human embryonic stem cell erythropoiesis by remodeling gene expression dynamics. *Epigenomics* 9, 1543–1558. doi: 10.2217/epi-2017-0039

Long, M. D., Smiraglia, D. J., and Campbell, M. J. (2017). The genomic impact of DNA CpG methylation on gene expression; relationships in prostate cancer. *Biomolecules* 7:E15. doi: 10.3390/biom7010015

Marques, C. J., Francisco, T., Sousa, S., Carvalho, F., Barros, A., and Sousa, M. (2010). Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil. Steril.* 94, 585–594. doi: 10.1016/j.fertnstert.2009.02.051

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Dsouza, C., Fouse, S. D., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253–257. doi: 10.1038/nature09165

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877. doi: 10.1093/nar/gki901

Moe, M., Meuwissen, T., Lien, S., Bendixen, C., Wang, X., Conley, L. C., et al. (2007). Gene expression profiles in testis of pigs with extreme high and low levels of androstenone. *BMC Genomics* 8:405. doi: 10.1186/1471-2164-8-405

Niemitz, E. L., and Feinberg, A. P. (2004). Epigenetics and assisted reproductive technology: a call for investigation. *Am. J. Hum. Genet.* 74, 599–609. doi: 10.1086/382897

Niesen, M. I., Osborne, A. R., Yang, H., Rastogi, S., Chellappan, S., Cheng, J. Q., et al. (2005). Activation of a methylated promoter mediated by a sequence-specific DNA-binding protein, RFX. *J. Biol. Chem.* 280, 38914–38922. doi: 10.1074/jbc.M504633200

Onteru, S. K., Fan, B., Du, Z. Q., Garrick, D. J., Stalder, K. J., and Rothschild, M. F. (2012). A whole-genome association study for pig reproductive traits. *Anim. Genet.* 43, 18–26. doi: 10.1111/j.1365-2052.2011.02213.x

Poplinski, A., Tüttelmann, F., Kanber, D., Horsthemke, B., and Gromoll, J. (2010). Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int. J. Androl.* 33, 642–649. doi: 10.1111/j.1365-2605.2009.01000.x

Rajender, S., Avery, K., and Agarwal, A. (2011). Epigenetics, spermatogenesis and male infertility. *Mutat. Res. Rev. Mutat. Res.* 727, 62–71. doi: 10.1016/j.mrrev.2011.04.002

Raza, M. A., Yu, N., Wang, D., Cao, L., Gan, S., and Chen, L. (2017). Differential DNA methylation and gene expression in reciprocal hybrids between *Solanum*

lycopersicum and *S. pimpinellifolium*. *DNA Res.* 24, 597–607. doi: 10.1093/dnares/dsx028

Razin, A., and Cedar, H. (1991). DNA methylation and gene expression. *Microbiol. Rev.* 55, 451–458. doi: 10.1002/wsbm.64

Saxonov, S., Berg, P., and Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1412–1417. doi: 10.1073/pnas.0510310103

Schachtschneider, K. M., Madsen, O., Park, C., Rund, L. A., Groenen, M. A. M., and Schook, L. B. (2015). Adult porcine genome-wide DNA methylation patterns support pigs as a biomedical model. *BMC Genomics* 16:743. doi: 10.1186/s12864-015-1938-x

Smith, Z. D., Gu, H., Bock, C., Gnirke, A., and Meissner, A. (2009). High-throughput bisulfite sequencing in mammalian genomes. *Methods* 48, 226–232. doi: 10.1016/j.ymeth.2009.05.003

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9440–9445. doi: 10.1073/pnas.1530509100

Strathe, A. B., Velander, I. H., Mark, T., and Kadarmideen, H. N. (2013). Genetic parameters for androstenone and skatole as indicators of boar taint and their relationship to production and litter size traits in Danish Landrace. *J. Anim. Sci.* 91, 2587–2595. doi: 10.2527/jas.2012-6107

Stuppia, L., Franzago, M., Ballerini, P., Gatta, V., and Antonucci, I. (2015). Epigenetics and male reproduction: the consequences of paternal lifestyle on fertility, embryo development, and children lifetime health. *Clin. Epigenet.* 7:120. doi: 10.1186/s13148-015-0155-4

Sun, Z., Cunningham, J., Slager, S., and Kocher, J. P. (2015). Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. *Epigenomics* 7, 813–828. doi: 10.2217/epi.15.21

Takai, D. (2001). Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum. Mol. Genet.* 10, 2619–2626. doi: 10.1093/hmg/10.23.2619

Verardo, L. L., Silva, F. F., Lopes, M. S., Madsen, O., Bastiaansen, J. W. M., Knol, E. F., et al. (2016). Revealing new candidate genes for reproductive traits in pigs: combining bayesian GWAS and functional pathways. *Genet. Sel. Evol.* 48:9. doi: 10.1186/s12711-016-0189-x

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* 31, 2912–2914. doi: 10.1093/bioinformatics/btv300

Wang, B., Li, L., Ni, F., Song, J., Wang, J., Mu, Y., et al. (2009). Mutational analysis of SAL-Like 4 (SALL4) in Han Chinese women with premature ovarian failure. *Mol. Hum. Reprod.* 15, 557–562. doi: 10.1093/molehr/gap046

Wang, H., Wang, J., Ning, C., Zheng, X., Fu, J., Wang, A., et al. (2017). Genome-wide DNA methylation and transcriptome analyses reveal genes involved in immune responses of pig peripheral blood mononuclear cells to poly I:C. *Sci. Rep.* 7:9707. doi: 10.1038/s41598-017-10648-9

Worm Ørntoft, M. B., Jensen, S. Ø, Hansen, T. B., Bramsen, J. B., and Andersen, C. L. (2017). Comparative analysis of 12 different kits for bisulfite conversion of circulating cell-free DNA. *Epigenetics* 12, 626–636. doi: 10.1080/15592294.2017.1334024

Wu, C., Ding, X., Li, H., Zhu, C., and Xiong, C. (2013). Genome-wide promoter methylation profile of human testis and epididymis: Identified from cell-free seminal DNA. *BMC Genomics* 14:288. doi: 10.1186/1471-2164-14-288

Wu, C. T., and Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* 293, 1103–1105. doi: 10.1126/science.293.5532.1103

Yang, X., Han, H., DeCarvalho, D. D., Lay, F. D., Jones, P. A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26, 577–590. doi: 10.1016/j.ccr.2014.07.028

Yuan, X. L., Gao, N., Xing, Y., Zhang, H. B., Zhang, A. L., Liu, J., et al. (2016). Profiling the genome-wide DNA methylation pattern of porcine ovaries using reduced representation bisulfite sequencing. *Sci. Rep.* 6:22138. doi: 10.1038/srep22138

Yuan, X. L., Zhang, Z., Li, B., Gao, N., Zhang, H., Sangild, P. T., et al. (2017). Genome-wide DNA methylation analysis of the porcine hypothalamus-pituitary-ovary axis. *Sci. Rep.* 7:4277. doi: 10.1038/s41598-017-04603-x

Zeitoun, K., Takayama, K., Sasano, H., Suzuki, T., Moghrabi, N., Andersson, S., et al. (1998). Deficient 17β-hydroxysteroid dehydrogenase type 2 expression in endometriosis: failure to metabolize 17β-estradiol. *J. Clin. Endocrinol. Metab.* 83, 4474–4480. doi: 10.1210/jc.83.12.4474

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* 14:244. doi: 10.1186/1471-2105-14-244

Zhang, L., Gu, C., Yang, L., Tang, F., and Gao, Y. Q. (2017). The sequence preference of DNA methylation variation in mammalians. *PLoS One* 12:e0186559. doi: 10.1371/journal.pone.0186559

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., et al. (2016). Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nat. Commun.* 7:12464. doi: 10.1038/ncomms12464

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang and Kadarmideen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 May 2019
doi: 10.3389/fgene.2019.00409

Dynamic Transcriptome Analysis Reveals Potential Long Non-coding RNAs Governing Postnatal Pineal Development in Pig

Yalan Yang[1†], Rong Zhou[2†], Wentong Li[1,2], Ying Liu[2], Yanmin Zhang[2], Hong Ao[2], Hua Li[1] and Kui Li[1,2]*

[1] Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, School of Life Sciences and Engineering, Foshan University, Foshan, China, [2] Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:
Yun Xiao,
Harbin Medical University, China
Tao Zhou,
Auburn University, United States

***Correspondence:**
Kui Li
likui@caas.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 28 January 2019
Accepted: 15 April 2019
Published: 03 May 2019

Citation:
Yang Y, Zhou R, Li W, Liu Y, Zhang Y, Ao H, Li H and Li K (2019) Dynamic Transcriptome Analysis Reveals Potential Long Non-coding RNAs Governing Postnatal Pineal Development in Pig. Front. Genet. 10:409. doi: 10.3389/fgene.2019.00409

Postnatal development and maturation of pineal gland is a highly dynamic period of tissue remodeling and phenotype maintenance, which is genetically controlled by programmed gene expression regulations. However, limited molecular characterization, particularly regarding long noncoding RNAs (lncRNA), is available for postnatal pineal at a whole transcriptome level. The present study first characterized the comprehensive pineal transcriptome profiles using strand-specific RNA-seq to illustrate the dynamic mRNA/lncRNA expression at three developmental stages (infancy, puberty, and adulthood). The results showed that 21,448 mRNAs and 8,166 novel lncRNAs were expressed in pig postnatal pineal gland. Among these genes, 3,573 mRNAs and 851 lncRNAs, including the 5-hydroxytryptamine receptors, exhibited significant dynamic regulation along maturation process, while the expression of homeobox genes didn't show significant differences. Gene Ontology analysis revealed that the differentially expressed genes (DEGs) were significantly enriched in ion transport and synaptic transmission, highlighting the critical role of calcium signaling in postnatal pineal development. Additionally, co-expression analysis revealed the DEGs could be grouped into 12 clusters with distinct expression patterns. Many differential lncRNAs were functionally enriched in co-expressed clusters of genes related to ion transport, transcription regulation, DNA binding, and visual perception. Our study first provided an overview of postnatal pineal transcriptome dynamics in pig and demonstrated that dynamic lncRNA regulation of developmental transitions impact pineal physiology.

Keywords: pineal gland, pig, long noncoding RNA, postnatal development, transcriptome

Abbreviations: DAVID, database for annotation, visualization, and integrated discovery; DGE, differentially expressed gene; GO, gene ontology; lincRNA, long intergenic noncoding RNAs; lncRNA, long noncoding RNAs; MCL, Markov clustering; PCA, principal component analysis; RPKM, reads per kilobase per million reads; RT-qPCR, real-time quantitative PCR; TF, transcription factor; Y, Yorkshire.

INTRODUCTION

The mammalian pineal gland is a neuroendocrine transducer whose main and most conserved function is converting photoperiodic information into the nocturnal hormonal signal of melatonin synthesis and secretion (Maronde and Stehle, 2007). Melatonin regulates a variety of circadian and circannual physiological processes, such as the sleep-wake cycle, feeding, and cognition rhythms (Leon et al., 2004; Acuna-Castroviejo et al., 2007). Recent studies have revealed that melatonin also regulates many general physiological functions, including lipid and glucose metabolism, immune function, and carcinogenesis (Carrillo-Vico et al., 2005; Jha et al., 2015; Trivedi et al., 2016). Exploring pineal development will contribute to an improved understanding of its functions and mechanisms of regulation. The pineal gland develops as a tubular evagination from the dorsal diencephalon between the habenular and posterior commissures in the embryonic brain. The pineal gland displays a phase of rapid cell proliferation during the prenatal periods. However, cell proliferation activity terminates rapidly (Sapède and Cau, 2013), and pinealoblasts differentiate into pinealocytes during the two first postnatal weeks in rats (Calvo and Boya, 1983). After postnatal maturation, the parenchyma of the pineal gland is composed primarily of pinealocytes and interstitial cells (Moller and Baeres, 2002).

Pineal development is a complicated and dynamic process that is precisely genetically controlled by the programmed expression of gene cascades and TFs. Several TFs responsible for the establishment and maintenance of the pineal phenotype have been identified, such as the homeobox TFs *PAX6*, *LHX9*, and *OTX2* (Rath et al., 2013). However, gene abundance represents only part of the complexity of the transcriptome, as it has emerged that lncRNAs, which are a subgroup of transcripts that are longer than 200 nucleotides (nt) yet have limited protein-coding potential, have recently emerged as pivotal regulators in governing various developmental processes (Batista and Chang, 2013). For example, lncRNAs could regulate skeletal muscle differentiation during myogenesis, such as MyoD and H19 (Dey et al., 2014; Gong et al., 2015a). LncRNAs show precisely spatiotemporal expression patterns and regulate specific neuronal functions in brain (Briggs et al., 2015). Until now, the expression dynamics of mRNAs and lncRNAs involved in pineal gland have not been extensively explored. Describing the transcriptome profiles of the pineal gland through development may improve our understanding of the molecular pathways and regulatory mechanisms that are responsible for postnatal pineal development in mammals.

The pig (*Sus scrofa*) not only is an important agricultural animal but also serves as an attractive model organism for biomedical research, due to the similarity of its organ size, anatomy and physiology, and developmental processes with those of humans (Groenen et al., 2012; Prather, 2013; Niu et al., 2017; Yan et al., 2018). Hence, we can understand the developmental patterns of the pineal gland in mammals by using the pig as a model. In this study, we characterized high-resolution pineal transcriptome profiles in Y pigs using strand-specific total RNA sequencing, which allowed us to comprehensively illustrate the dynamic characteristics and functions of mRNAs/lncRNAs across three postnatal developmental stages: infancy (30 days, Y30), puberty (180 days, Y180), and adulthood (300 days, Y300). These results establish a general overview of the pineal transcriptome dynamics and pave the road for further investigations of the underlying functions and regulatory mechanisms of lncRNAs governing postnatal development of the mammalian pineal gland.

MATERIALS AND METHODS

Sample Collection

Nine Y pigs with the same genetic background at postnatal days 30, 180, and 300 (three replicates per stage) were obtained from the Tianjin Ningheyuan Swine Breeding Farm (Tianjin, China) and slaughtered during daytime, between 10:00 and 14:00 Beijing time. The pineal sample of each pig was collected and immediately frozen in liquid nitrogen until RNA isolation. All animal procedures were performed according to the protocols of the Chinese Academy of Agricultural Sciences and the Institutional Animal Care and Use Committee.

Transcriptome Library Preparation and Sequencing

Total RNA from pineal glands was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's directions. The purified RNA was treated with DNase I (Qiagen, Beijing, China). The quantity and purity of the RNA samples were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, United States). Ribosomal RNA was depleted using the Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, Madison, WI, United States). Next, strand-specific RNA-seq libraries for paired-end sequencing were prepared using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, United States) according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 4000 platform to generate 150 bp paired-end reads (Novogene Bioinformatics Technology Co. Ltd., Tianjin, China).

Transcriptome Assembly

The raw reads were firstly subjected to remove adaptor sequences and low-quality reads using custom scripts. The processed clean reads from each sample were then mapped to the reference genome of *Sus scrofa* (v11.1) using TopHat2 (v2.1.0) (Trapnell et al., 2009) with known gene annotation, parameters were set for strand-specific mapping (library-type "fr-secondstrand"). The reference genome sequence and gene annotation files were downloaded from the Ensembl database (release 90)[1]. After mapping, duplicate reads were removed using the rmdup tool in the samtools package (Li et al., 2009) to limit the influence of PCR artifacts. The remaining unique mapped reads of each sample were assembled into transcripts independently using Cufflinks (v1.3.0) (Trapnell et al., 2012) with the assistance of known annotations. Finally, assembled transcripts from each sample

[1]http://asia.ensembl.org/index.html

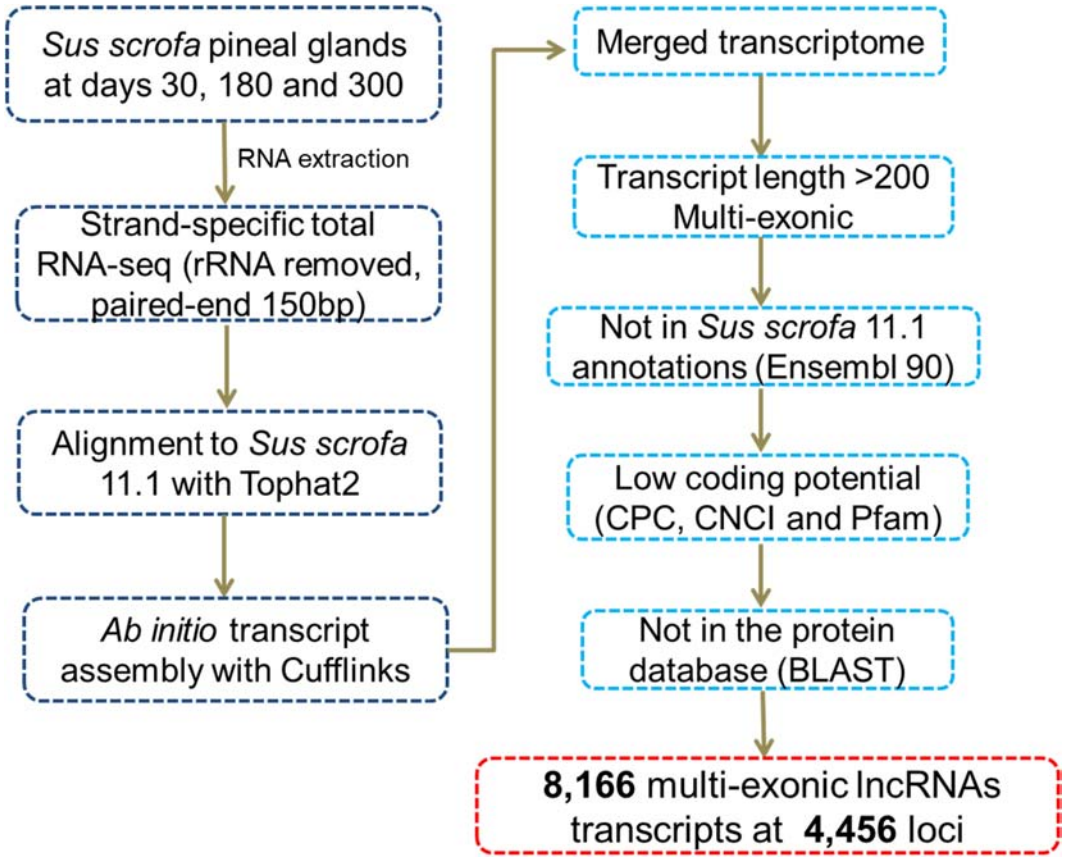


FIGURE 1 | Pipeline for the identification of novel lncRNAs.

were merged into a consensus transcriptome using Cuffmerge (v1.0.0) (Trapnell et al., 2012).

Identification of lncRNAs

We identified novel lncRNAs in the pig pineal transcriptome using similar methods to those reported in our previous studies (Tang et al., 2017; Yang et al., 2017). A series of stringent filtering steps were utilized (**Figure 1**) as follows: (i) Single-exon transcripts and the transcripts less than 200 bp were removed to avoid unreliable transcripts; (ii) We filtered transcripts overlapping (>1 bp) with known gene models deposited in the Ensembl database; (iii) Coding Potential Calculator (CPC, v0.9-r2) (Kong et al., 2007) and Coding-Non-Coding Index (CNCI, v2) (Sun et al., 2013) programs were used to evaluate the coding potential of each transcript. Transcripts predicted to have coding potential (score > 0) by any of these two programs were filtered out; (iv) The transcripts whose corresponding translated protein sequences had a known protein-coding domain in the Pfam database (version 30.0) were removed by PfamScan (v1.3) (Finn et al., 2014); and (v) BLASTs (BLAST 2.2.26+) was used to remove transcripts with similarity to known proteins in the UniRef90 database (UniProt Consortium, 2015) with an *E*-value cutoff of 10^{-5}. Transcripts remaining after the stringent filtering described above were considered putative lncRNAs.

Expression Analysis

The raw read counts for each gene (mRNA/lncRNA) were calculated using HTSeq-count (Anders et al., 2015). For genes with multiple transcripts of different lengths, the longest transcript was selected to compute the gene expression level, measured as RPKM. Genes with RPKM $\geq$ 0.1 in at least one sample were defined as expressed genes. Highly expressed genes were defined as genes with a maximum RPKM $\geq$ 50 across the samples (Wang et al., 2014). The edgeR (exact test for negative binomial distribution) Bioconductor package (Robinson et al., 2009) in R software was used to identify DEGs between developmental stages. Gene expression normalization among samples to adjust for different sequencing depths across samples was performed using edgeR (Robinson et al., 2009). After estimating the dispersion of each gene, significantly DEGs were identified using cutoffs of false discovery rate (FDR) $\leq$ 0.05 and $|\log2\ FC| \geq 1$ according to the edgeR's recommendation (Robinson et al., 2009) and previous studies (Xue et al., 2013; Vanlandewijck et al., 2018).

Co-expression Network Construction

Normalized, non-log transformed gene expression data (RPKM values) of all the differentially expressed mRNAs/lncRNAs were imported into Biolayout Express (3D) (Theocharidis et al., 2009).

A pairwise gene-to-gene Pearson correlation matrix was calculated as a measure of similarity between genes. Based on a Pearson correlation coefficient cut-off threshold of $r \geq 0.90$, a weighted, undirected co-expression network of mRNA-lncRNA interactions was generated. In this network, each node represents one gene (mRNA or lncRNA) and the edge between two nodes represents the Pearson correlation coefficients above the selected threshold. The network was clustered into groups of mRNAs/lncRNAs sharing similar expression patterns using the MCL algorithm (Enright et al., 2002), which has been demonstrated to be one of the most effective graph-based clustering algorithms available. To control the size of the clusters, the inflation coefficient was set to 2.4 and each cluster must contain at least 30 genes. This network was checked manually and clusters with no particular expression pattern were removed. Clusters were named according to their relative size, the largest cluster being designated Cluster 1.

Functional Enrichment Analysis

Gene ontology enrichment analysis were performed by the DAVID website (v6.7[2]) (Huang et al., 2008) with a background set of human orthologues included in this study.

Real-Time Quantitative PCR (RT-qPCR)

The total RNA of each sample was reverse transcribed into cDNA using a RevertAid First Strand cDNA Synthesis Kit (Thermo, Waltham, MA, United States) according to the manufacturer's instructions. The RT-qPCR reaction solution was comprised of 10 μl of 2× SYBR Premix Ex Taq (Takara, Dalian, China), 0.4 μl of each primer, 1 μl of cDNA, 0.4 μl of Dye II, and sterile water to a volume 20 μl. The RT-qPCR cycling parameters were as follows: 95°C for 5 min, followed by 40 cycles at 95°C for 5 s and 60°C for 1 min. Next, a dissociation program was carried out at 95°C for 15 s, 60°C for 1 min, and 95°C for 15 s. Each reaction was performed in triplicate. The 2-ΔΔCt method was used to determine the gene expression level. The porcine *GAPDH* gene was selected as an internal control. All primer sequences are listed in **Supplementary Table S1**.

[2]http://david.abcc.ncifcrf.gov/

RESULTS

Overview of the *Sus scrofa* Pineal Transcriptome Data

To identify changes in mRNA/lncRNA expression during postnatal pineal gland development, we generated RNA-seq libraries from the pineal glands of female Y pigs at infancy (Y30), puberty (Y180), and adulthood (Y300). Three biological replicates were evaluated per stage. Utilizing strand-specific RNA-seq of total RNA, a total of 1.05 billion clean sequencing reads (150 bp paired-end) were obtained after discarding low-quality and adaptor reads, corresponding to an average of 116.4 million sequence reads per sample. Of the clean reads, 79.9–90.0% could be mapped to the pig reference genome (version 11.1) by the Tophat2 pipeline (Trapnell et al., 2009) (**Table 1**). After removing duplicate reads, the remaining uniquely mapped reads were used for further lncRNA identification and gene expression analyses.

Identification and Characterization of lncRNAs in the Pineal Transcriptome

After reconstructing the transcriptome using Cufflinks and Cuffmerge (Trapnell et al., 2009), we identified putative lncRNAs in the pineal transcriptome using a pipeline (**Figure 1**) similar to those reported in our previous studies (Tang et al., 2017; Yang et al., 2017). Eventually, a total of 8,166 multi-exonic lncRNA transcripts corresponding to 4,456 genomic loci were obtained (**Supplementary Table S2**). According to their genomic location, most (6,505) were lincRNAs located in intergenic regions, while 1,129 were lncRNAs transcribed from the antisense strand of the reference coding transcript, and the remaining 532 lncRNAs overlapped with middle coding exon regions.

We next analyzed the features of these newly identified lncRNAs, namely novel lncRNAs. As expected, the novel lncRNAs contained fewer exons (3.1 exons on average) than mRNAs (11.6 exons on average; $P < 2.2e-16$) (**Figure 2A**). The average transcript length of these lncRNAs (2235.2 nt) was significantly shorter than that of mRNAs (3296.1 nt; $P < 2.2e-16$) (**Figure 2B**). Moreover, the expression levels of the lncRNAs (average RPKM = 2.7) were also significantly lower than those of the mRNAs (average RPKM = 10.9; $P < 2.2e-16$)

TABLE 1 | Summary of sequencing metrics and read mapping for the RNA-seq of pig pineal glands.

Sample	Stage	Length	Reads	Mapped reads		Mapped ratio	
				Read1	Read2	Read1	Read2
Y30_1	30 days	150 bp	54,049,538 × 2	47,598,866	43,352,218	88.1%	80.2%
Y30_2	30 days	150 bp	50,796,645 × 2	44,735,289	41,004,165	88.1%	80.7%
Y30_3	30 days	150 bp	58,589,821 × 2	51,633,312	46,824,908	88.1%	79.9%
Y180_1	180 days	150 bp	79,812,929 × 2	70,250,218	67,703,427	88.0%	84.8%
Y180_2	180 days	150 bp	55,857,541 × 2	49,638,178	45,750,849	88.9%	81.9%
Y180_3	180 days	150 bp	58,459,709 × 2	51,560,101	46,995,972	88.2%	80.4%
Y300_1	300 days	150 bp	58,236,109 × 2	51,208,023	46,802,188	87.9%	80.4%
Y300_2	300 days	150 bp	53,832,028 × 2	48,429,903	44,651,018	90.0%	82.9%
Y300_3	300 days	150 bp	54,051,461 × 2	48,432,224	43,497,472	89.6%	80.5%

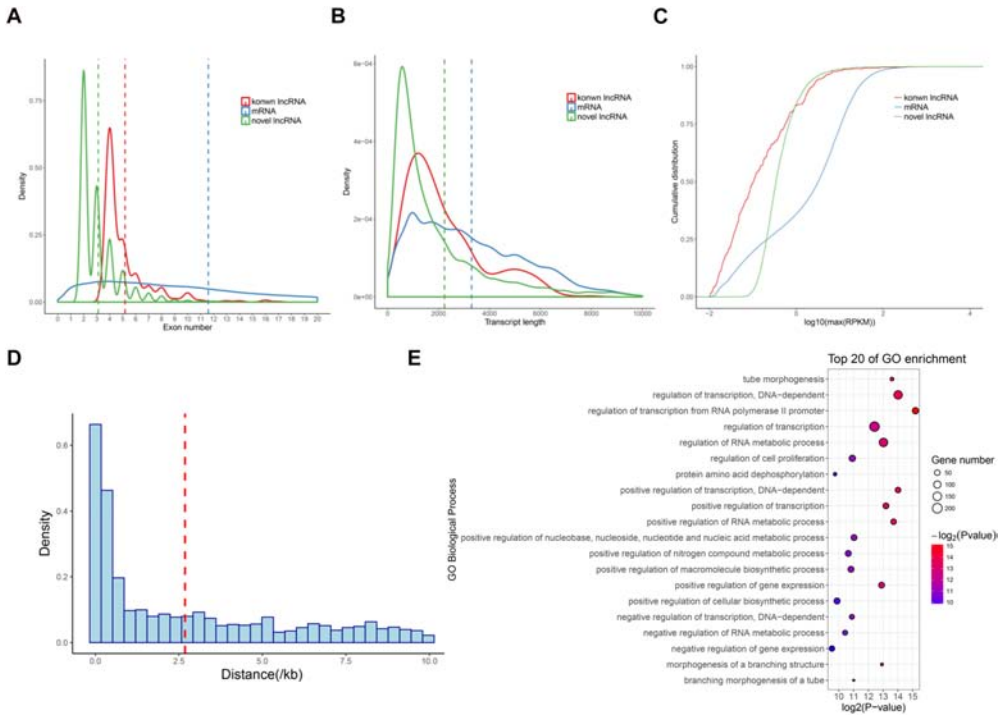


FIGURE 2 | Characterization of lncRNAs in pig pineal glands. **(A)** Comparison of exon number between lncRNAs and mRNAs. **(B)** Comparison of transcript length between lncRNAs and mRNAs. **(C)** Comparison of expression level between lncRNAs and mRNAs. **(D)** The distribution of the distance from lincRNAs to their nearest neighboring protein-coding genes. The average (red dashed line) distance is indicated. **(E)** GO biological processes analysis of the neighboring protein-coding genes of the lincRNAs.

(**Figure 2C**). These results were consistent with the previous lncRNAs reports in pigs and other mammals (Iyer et al., 2015; Tang et al., 2017; Yang et al., 2017). Additionally, we found that 2,481 lincRNA were transcribed near (<10 kb) their protein-coding neighbors. The average distance from lincRNAs to their neighboring genes was 2.68 kb (**Figure 2D**). GO analysis revealed that these neighboring genes were significantly enriched in regulation of transcription and tube morphogenesis functions (**Figure 2E**), indicating that these lincRNAs are preferentially located in the vicinity of genes with specific functions that are closely associated with postnatal pineal development.

Dynamic Expression of mRNAs and lncRNAs in Pineal Gland

We next evaluated the expression of novel lncRNAs, known lincRNAs, and mRNAs across postnatal pineal development and found a high Pearson correlation within and across stages ($R > 0.95$) (**Figure 3A**), indicating a high level of measurement consistency among biological replicates. A PCA was performed in order to understand the expression patterns of all mRNAs and lncRNAs during postnatal pineal development. We found that the PCA could clearly separate the three developmental stages from each other; the first two principal components (PC1 and PC2) could explain 35.1 and 23.5% of the transcriptional variation, respectively (**Figure 3B**). Clustering analysis revealed that samples within stages were clustered together first, and then Y30 and Y180 were grouped to form a larger cluster, and finally, clustered with Y300 (**Figure 3C**). These findings demonstrated a very high reproducibility within stages and distinct expression patterns across postnatal pineal development.

We detected an average of 15,388 mRNAs (a total of 21,448 mRNAs, with a range of 15,227–15,611 mRNAs per sample) and 2,740 lncRNAs (2,511–3,015 lncRNAs per sample) expressed (RPKM $\geq$ 0.1) in pineal glands, which accounted for 68.9 and 57.0% of the total mRNAs and lncRNAs, respectively. The RPKM values of most of the mRNAs were greater than 1, while the majority of the lncRNAs were lowly expressed (RPKM $\leq$ 0.1). Of these RNAs, 853 genes (842 mRNAs and 11 lncRNAs) were highly expressed in pineal glands (RPKM $\geq$ 50 in at least one sample). As expected, these genes were significantly enriched in translation, oxidative phosphorylation, and ATP synthesis-coupled electron transport functions (**Figure 3D**), all of which are essential for protein synthesis and other basic requirements for postnatal pineal development. Additionally, *TTR*, a pineal-specific gene, was highly expressed in our samples, especially at the Y30 stage (**Figure 3E**). Most of the homeobox TFs were lowly expressed in postnatal pineal gland (**Figure 3F**).

Differentially Expressed mRNAs and lncRNAs

We found a total of 4,424 genes (including 3,573 mRNAs and 851 lncRNAs) with a significant difference in expression ($|\log_2$

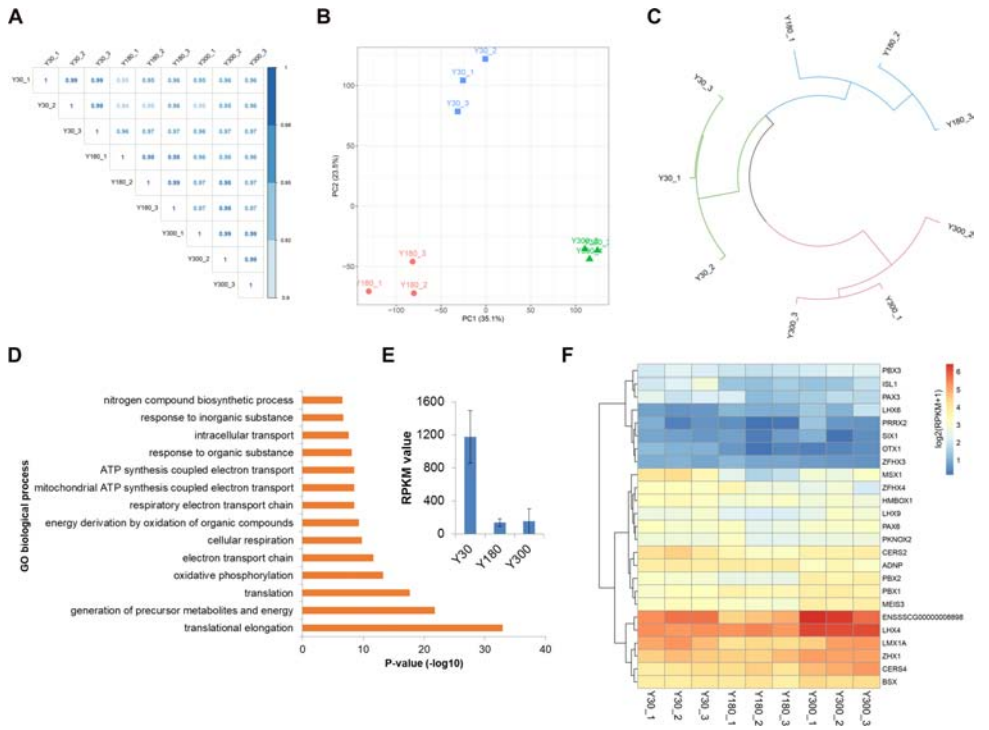


FIGURE 3 | Dynamic expression profiles of mRNAs and lncRNAs during porcine postnatal pineal development. **(A)** Pearson correlation plot for the pineal transcriptome of Y30, Y180, and Y300, with three replicates for each developmental stage. **(B)** Principal component analysis of the pineal samples across three postnatal developmental stages based on both mRNA and lncRNA expression levels. Stages are illustrated by different shapes and colors. The *x*- and *y*-axes represent the first and second PC, respectively, with the percent variance explained by each PC in parentheses. **(C)** Hierarchical clustering analysis of the nine pineal samples across three developmental stages based on both mRNA and lncRNA expression levels. **(D)** Top enriched GO biological process terms of the highly expressed genes in the pineal gland. **(E)** The expression of *TTR* gene in postnatal pineal gland. **(F)** Heatmap showing the expression of homeobox transcription factors during postnatal pineal development.

fold change (FC)|≥1 and FDR≤0.05) between developmental stages, including 2,417 Y180-Y30 (including 1,982 mRNAs and 436 lncRNAs), 2,788 Y300-Y30 (including 2,264 mRNAs and 524 lncRNAs), and 1,633 Y300-Y180 (including 1,187 mRNAs and 446 lncRNAs) DEGs (**Figures 4A,B**). Several 5-hydroxytryptamine (serotonin) receptors were included in this list, including *HTR2A*, *HTR2B*, *HTR2C*, and *HTR7*. We randomly verified 15 of the DEGs (10 mRNAs and 5 lncRNAs) by RT-qPCR and found a high concordance between the RT-qPCR and the RNA-seq data (**Figure 4C**), suggesting that the differential expression analysis based on the RNA-seq data was reliable. The highest number of DEGs was observed in the Y300-Y30 comparison, which was correlated with the difference in development time among the three stages. Most DEGs were observed in at least two of the three comparisons, and 91 of them (56 mRNAs and 35 lncRNAs) were found in all three comparisons (**Figure 4D**), including genes related to phosphate metabolic (*PPM1J*, *ND4*, *PRLR*, and *ND5*) and cell motility (*CCK*, *FOXJ1*, *DCDC2*, and *DNAH2*). We further examined the enriched functions of the DEGs through GO enrichment analysis. Compared with Y30, the up-regulated genes in Y180 were significantly enriched in ion transport, transmission of nerve impulse, cell-cell signaling, and synaptic transmission functions, while the down-regulated genes were associated with ion transport, oxidation-reduction, and cell cycle categories (**Figure 4E** and **Supplementary Table S3**). The up-regulated genes in Y300 when compared with Y30 were significantly enriched in sensory perception of light stimulus, visual perception, transmission of nerve impulse, and neurological system process functions, while the down-regulated genes were enriched in ion transport and cell cycle functions (**Figure 4F** and **Supplementary Table S3**). Compared with Y180, the up-regulated genes in Y300 were significantly enriched in mitochondrion organization, ribosome biogenesis, and negative regulation of cell cycle process functions, while the down-regulated genes were associated with transcription and RNA metabolic and phosphate metabolic processes (**Figure 4G** and **Supplementary Table S3**).

Inference of Pineal lncRNA Function Using Co-expressed Network

To explore the potential functions and regulatory mechanisms of lncRNAs during postnatal pineal development, we constructed a co-expression interaction network of differentially expressed mRNAs and lncRNAs. The network consisted of 605,831 interaction pairs. These genes were grouped into 12 co-expression clusters by MCL algorithm (Enright et al., 2002).

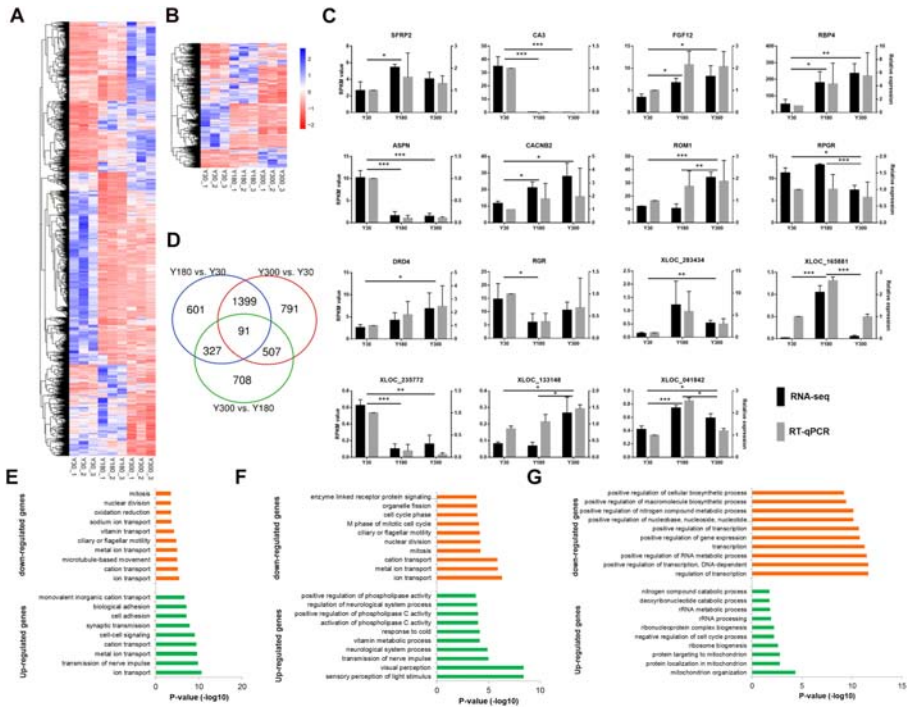


FIGURE 4 | Differential expression analysis of mRNAs and lncRNAs during porcine postnatal pineal development. **(A,B)** Heatmap showing the differentially expressed mRNAs **(A)** and lncRNAs **(B)** during porcine postnatal pineal development. **(C)** Experimental validation of RNA-seq data by RT-qPCR. Gene expression differences between developmental stages were evaluated based on the RNA-Seq data using the edgeR package. Error bars are SEM, $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. **(D)** Venn diagram showing the number of differentially expressed mRNAs and lncRNAs between different development stages. **(E–G)** GO biological process analysis of the up-regulated and down-regulated genes between Y180-Y30 **(E)**, Y300-Y30 **(F)**, and Y300-Y180 **(G)**.

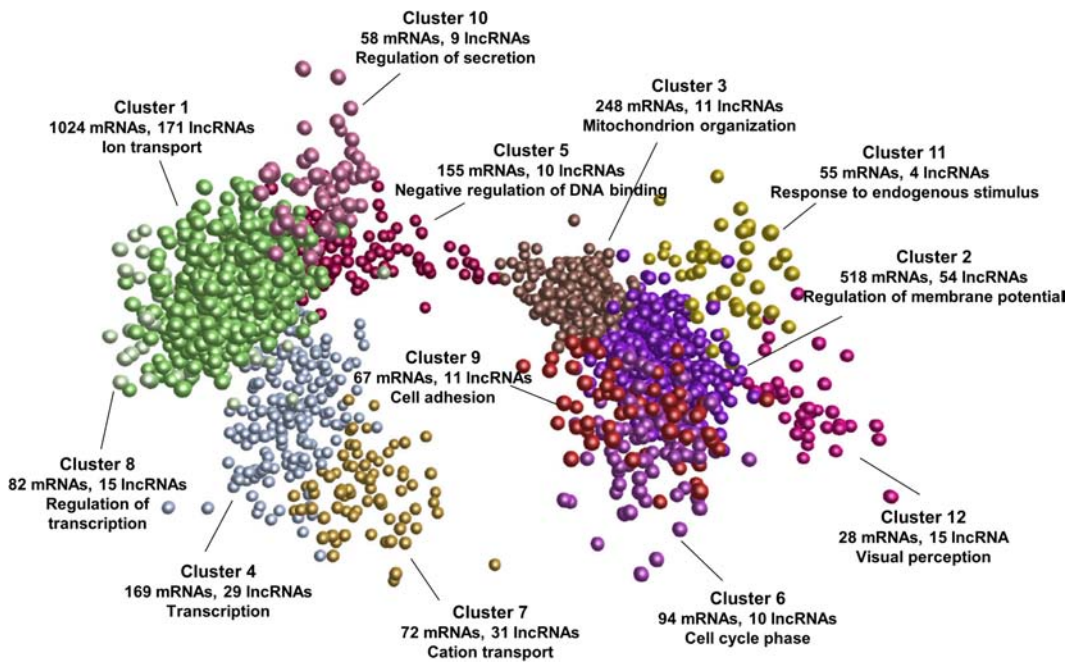


FIGURE 5 | mRNA-lncRNA co-expression network. All the differentially expressed mRNAs and lncRNAs were used to construct the co-expression network by Biolayout Express (3D). The number of mRNAs/lncRNAs in each co-expression cluster and the most significantly enriched GO biological process of each cluster were shown.

The expression pattern of each cluster during postnatal pineal development was shown in **Supplementary Figure S1**. Some of these clusters contained mRNAs that are closely associated with postnatal pineal development (**Figure 5** and **Supplementary Table S4**). Cluster 1 was the biggest one, which contained 1024 mRNAs and 171 lncRNAs, genes in this cluster were highly expressed at Y30 stage, such as members of the solute carrier family genes (*SLC5A5*, *SLC13A5*, and *SLC39A12*). Cluster 7 was highly expressed at Y180 stage and contained 72 mRNAs and 31 lncRNAs. Interestingly, GO enrichment analyses suggested that ion transport was the most significantly enriched term of genes in these two clusters. The genes in both cluster 2 (518 mRNAs and 54 lncRNAs) and cluster 3 (248 mRNAs and 11 lncRNAs) were abundantly expressed at Y300 stage. The genes in cluster 3 were higher expressed at Y30 than at Y180 stage, while the genes in cluster 2 were stably expressed at these two stages. Cluster 2 and cluster 3 mainly functioned in regulation of membrane potential and mitochondrion organization, respectively. Additionally, cluster 4 (169 mRNAs and 29 lncRNAs) was enriched with transcription and oxidative phosphorylation genes, including the core subunits of mitochondrial membrane respiratory chain NADH dehydrogenase (*ND1*, *ND2*, *ND4*, and *ND5*). The genes in cluster 4 were higher expressed at Y180 stage than at Y30 and Y300 stages. Whereas the genes in cluster 5 exhibited an inversed expression patterns with the genes in cluster 4. Negative regulation of DNA binding was the most enriched biological process for cluster 5, which contained 155 mRNAs and 10 lncRNAs, such as *PTHLH*, *SMO*, *ID1*, and *XLOC_050558*. Genes in cluster 6 (94 mRNAs and 10 lncRNAs) were specifically expressed at Y300 stage, which were closely associated with cell cycle phase and mitosis, such as *CCNB1*, *CDC20*, and *DLGAP5*. Remarkably, the expressions of genes in cluster 9 (67 mRNAs and 11 lncRNAs) and cluster 10 (58 mRNAs and 9 lncRNAs) were continuously increased during postnatal pineal development, cell adhesion and regulation of secretion was the most significantly enriched biological processes in these two clusters, respectively. The continuously decreased genes were grouped into cluster 8 (82 mRNAs, 15 lncRNAs) and mostly enriched in regulation of transcription, such as *PLAG1*, *ATF7IP*, and *CRTC3*. Genes in cluster 11 and cluster 12 were associated with response to endogenous stimulus and visual perception, respectively. These results suggested putative regulatory functions for a subset of lncRNAs in postnatal pineal development.

DISCUSSION

In this study, we provided deep strand-specific RNA-seq of total RNA from three representative postnatal developing stages (infancy, puberty, and adulthood) of the porcine pineal gland, which we used to study the expression profiles of mRNAs/lncRNAs. We expect that this new resource will contribute to the understanding of the importance of transcription regulation in mammalian postnatal pineal gland development and maturation.

Pineal gland is an neuroendocrine organ for the regulation of the circadian clock system in all vertebrate species (Macchi and Bruce, 2004). It's well known that homeobox genes are essential for normal pineal development and are key regulators in the maintenance of the postnatal pineal phenotype (Rath et al., 2013). We observed that most of the homeobox genes were lowly expressed (RPKM < 1) in our study, *HOPX* and *LHX4* were the most abundantly expressed ones in pig postnatal pineal gland, implying these two genes might play important roles in pineal development, though their function in pineal has not been reported. *LHX9* and *PAX6* are essential for early development of the mammalian pineal gland (Rath et al., 2013; Yamazaki et al., 2015), our study confirmed that these two genes were lowly expressed in the postnatal pineal gland after 30 days. *OTX2* displayed decreased expression in the postnatal pineal gland of rat (Rath et al., 2006), and was barely expressed in our samples. Additionally, the neurogenic differentiation factor 1 (*NEUROD1*) gene, a member of the bHLH TF family, is known to influence the fate of specific neuronal and endocrine cells (Munoz et al., 2007), and we confirmed that it was highly expressed in the postnatal pineal gland.

The expression of most DEGs changed constantly across postnatal pineal development, reflecting a dynamic regulation of gene expression. For example, the expression level of the transthyretin (*TTR*) gene, the major thyroid hormone transporter in the CNS, was much higher at Y30 than at Y180 or Y300. *TTR* has also been reported to be differentially expressed between midnight and midday in the pineal gland (Acuna-Castroviejo et al., 2007). We observed that DEGs were significantly associated with the ion transport, cell-cell signaling, synaptic transmission, and developmental maturation. Especially, 48 genes in calcium signaling pathway were differentially expressed, such as *CALB1* and *CACNB2*, which are involved in a variety of calcium-dependent processes, including cell motility, cell division, and hormone or neurotransmitter release (Dolphin, 2007). It's reported that melatonin could modulate neural development through the regulation of calcium signaling (Poloni et al., 2011). Additionally, 88 DEGs were involved in transmission of nerve impulse, such as *HTR2A*, *HTR2C*, and *HTR7*, which are 5-hydroxytryptamine (serotonin) receptors. 5-hydroxytryptamine is a precursor for melatonin production and is produced abundantly in the pineal gland of all vertebrate animals (Sapède and Cau, 2013). These results provide evidences that the critical roles of ion transport, especially calcium signaling, in postnatal pineal development, which might contribute to deeply understand the complexity of the pineal architectures and functions.

With the rapid adoption of RNA-seq technologies, thousands of lncRNA in the genome have been discovered in various species, their functions in various biological processes have been demonstrated (Geng et al., 2013; Gong et al., 2015b; Volders et al., 2015; Liang et al., 2018). However, compared with those of human and mouse, the lncRNA resources in pig are relatively limited (Quek et al., 2015; Liang et al., 2018). In this study, we identified a total of 8,166 novel lncRNAs, greatly expanding the genomic information of non-coding RNAs in pigs. These lncRNAs exhibited similar genomic characteristics with those of lncRNAs described in previous studies of pigs and other mammals (Iyer et al., 2015; Tang et al., 2017;

Yang et al., 2017). 851 lncRNAs, including 35 known and 816 novel lncRNAs, were differentially expressed across postnatal pineal development. Remarkably, 282 of them were transcribed near their protein-coding neighbors. For instance, XLOC_199747 located upstream of neurotrophin 3 (*NTF3*). There was a significantly positive correlation between the expressions of these two genes (r = 0.81). These results suggested that these differentially expressed lncRNAs might act on mRNAs involving in postnatal development by *cis* regulation. Co-expression analysis identified coordinated gene clusters that were shared in a developmental-specific expression fashion, which is an effective approach to uncover the function of lncRNAs (Pauli et al., 2012; Anamaria et al., 2014). We found most clusters containing genes with interesting functions. For example, GO enrichment analyses suggested that the cluster 1 was mostly associated with ion-transport, including many solute carrier family genes, which play important roles in the adrenergic regulation of cAMP and cGMP in pinealocytes (Sugden et al., 1986). Genes in cluster 2 were highly expressed at Y300 stages, which were closely related to regulation of membrane potential and transmission of nerve impulse, implying the critical roles of these mRNAs (such as *SCN1B* and *SYT4*) and lncRNAs (such as *XLOC_018250* and *XLOC_179558*) in mature pineal gland. Cell cycle genes (such as *CCNB1*, *CCNB2*, *CCNB3*, and *CCNF*), and 10 lncRNAs (such as *XLOC_280714* and *XLOC_156756*) were grouped into Cluster 6. The expression of these genes was decreased dramatically during postnatal pineal development, which was in consistence with the termination of pinealoblasts proliferation after birth (Sapède and Cau, 2013). Another intriguing example is cluster 12, which includes 15 lncRNAs (such as *XLOC_046348* and *XLOC_196944*). The mRNAs in this cluster were enriched in functional terms related to visual and sensory perception, which have been shown to play essential roles in circadian melatonin rhythm (Reiter, 1993). The dynamic changes observed in the co-expression networks offer insights regarding to the functions and regulation of lncRNAs during postnatal pineal development.

CONCLUSION

Overall, our data cataloged the pineal transcriptional profiles and basic gene expression features during postnatal development and maturation in pig. Novel lncRNAs were identified, which provide rich resources for understanding the molecular mechanisms and regulatory network of postnatal pineal development in mammals. The lncRNAs in the co-expression network may be considered as promising targets for postnatal pineal development, maturation, and phenotype maintenance, but their function still needs to be further explored at the molecular, cellular, and individual levels.

ETHICS STATEMENT

All animal procedures were performed according to the protocols of the Chinese Academy of Agricultural Sciences and the Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

KL designed and managed the project. YY administered the computational analysis. YY and RZ analyzed the data and wrote the manuscript. RZ, WL, YL, YZ, and YY performed animal work and collected biological samples. WL, YL, and YZ performed molecular experiments. HL, HA, and KL revised the manuscript. All the authors approved the final manuscript.

FUNDING

This work was supported by National Key Basic Research Program of China (2015CB943100), National Science and Technology Major Project (2018ZX08009-26B, 2016ZX080011-006), the Key Project of National Natural Science Foundation of China (31330074), National Nonprofit Institute Research Grant (Y2016JC07 and 2018-YWF-YB-7), Foshan University Initiative Scientific Research Program, and Open Subject of Key Laboratory of Animal Molecular Design and Precise Breeding (2018A09).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00409/full#supplementary-material

FIGURE S1 | The dynamic expression pattern of each co-expression cluster during postnatal pineal development.

TABLE S1 | Primer sequences of mRNAs and lncRNAs selected for validation by RT-qPCR.

TABLE S2 | Summary of the predicted lncRNAs in the *Sus scrofa* pineal gland.

TABLE S3 | GO biological process enrichment analysis of the differentially expressed genes between different developmental stages.

TABLE S4 | GO biological process enrichment analysis of the mRNAs in each co-expression cluster.

REFERENCES

Acuna-Castroviejo, D., Escames, G., Rodriguez, M. I., and Lopez, L. C. (2007). Melatonin role in the mitochondrial function. *Front. Biosci.* 12, 947–963.

Anamaria, N., Magali, S., Maria, W., Angélica, L., Tasman, D., Ulrich, Z., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Batista, P. J., and Chang, H. Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152, 1298–1307. doi: 10.1016/j.cell.2013.02.012

Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L., and Guy, B. (2015). Mechanisms of long non-coding rnas in mammalian nervous system

development, plasticity, disease, and evolution. *Neuron* 88, 861–877. doi: 10.1016/j.neuron.2015.09.045

Calvo, J., and Boya, J. (1983). Postnatal development of cell types in the rat pineal gland. *J. Anat.* 137(Pt 1), 185–195.

Carrillo-Vico, A., Guerrero, J. M., Lardone, P. J., and Reiter, R. J. (2005). A review of the multiple actions of melatonin on the immune system. *Endocrine* 27, 189–200. doi: 10.1385/endo:27:2:189

Dey, B. K., Pfeifer, K., and Dutta, A. (2014). The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev.* 28, 491–501. doi: 10.1101/gad.234419.113

Dolphin, A. (2007). Voltage-dependent calcium channels. *Encycl. Pain* 24(Suppl. 1):2654. doi: 10.1007/978-3-540-29805-2_4827

Enright, A. J., Dongen, S., and Van Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Geng, C., Wang, Z., Wang, D., Qiu, C., Liu, M., Xing, C., et al. (2013). LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 41, D983–D986. doi: 10.1093/nar/gks1099

Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S., et al. (2015a). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* 34, 181–191. doi: 10.1016/j.devcel.2015.05.009

Gong, J., Liu, W., Zhang, J., Miao, X., and Guo, A. Y. (2015b). lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.* 43, D181–D186. doi: 10.1093/nar/gku1000

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi: 10.1038/ng.3192

Jha, P. K., Challet, E., and Kalsbeek, A. (2015). Circadian rhythms in glucose and lipid metabolism in nocturnal and diurnal mammals. *Mol. Cell. Endocrinol.* 418(Pt 1), 74–88. doi: 10.1016/j.mce.2015.01.024

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35(Suppl. 2), W345–W349.

Leon, J., Acuna-Castroviejo, D., Sainz, R. M., Mayo, J. C., Tan, D. X., and Reiter, R. J. (2004). Melatonin and mitochondrial function. *Life Sci.* 75, 765–790. doi: 10.1016/j.lfs.2004.03.003

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(1 Pt 2), 1653–1654.

Liang, G., Yang, Y., Li, H., Yu, H., Li, X., Tang, Z., et al. (2018). LncRNAnet: a comprehensive *Sus scrofa* lncRNA database. *Anim. Genet.* 49, 632–635. doi: 10.1111/age.12720

Macchi, M. M., and Bruce, J. N. (2004). Human pineal physiology and functional significance of melatonin. *Front. Neuroendocrinol.* 25:177–195. doi: 10.1016/j.yfrne.2004.08.001

Maronde, E., and Stehle, J. H. (2007). The mammalian pineal gland: known facts, unknown facets. *Trends Endocrinol. Metab.* 18, 142–149. doi: 10.1016/j.tem.2007.03.001

Moller, M., and Baeres, F. M. M. (2002). The anatomy and innervation of the mammalian pineal gland. *Cell Tissue Res.* 309, 139–150. doi: 10.1007/s00441-002-0580-5

Munoz, E. M., Bailey, M. J., Rath, M. F., Shi, Q., Morin, F., Coon, S. L., et al. (2007). NeuroD1: developmental expression and regulated genes in the rodent pineal gland. *J. Neurochem.* 102, 887–899. doi: 10.1111/j.1471-4159.2007.04605.x

Niu, D., Wei, H. J., Lin, L., George, H., Wang, T., Lee, I. H., et al. (2017). Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* 357, 1303–1307. doi: 10.1126/science.aan4187

Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. doi: 10.1101/gr.133009.111

Poloni, J. D., Feltes, B. C., and Bonatto, D. (2011). Melatonin as a central molecule connecting neural development and calcium signaling. *Funct. Integrat. Genomics* 11, 383–388. doi: 10.1007/s10142-011-0221-8

Prather, R. S. (2013). Pig genomics for biomedicine. *Nat. Biotechnol.* 31, 122–124.

Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., et al. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43, D168–D173. doi: 10.1093/nar/gku988

Rath, M. F., Munoz, E., Ganguly, S., Morin, F., Shi, Q., Klein, D. C., et al. (2006). Expression of the Otx2 homeobox gene in the developing mammalian brain: embryonic and adult expression in the pineal gland. *J. Neurochem.* 97, 556–566. doi: 10.1111/j.1471-4159.2006.03773.x

Rath, M. F., Rohde, K., Klein, D. C., and Moller, M. (2013). Homeobox genes in the rodent pineal gland: roles in development and phenotype maintenance. *Neurochem. Res.* 38, 1100–1112. doi: 10.1007/s11064-012-0906-y

Reiter, R. J. (1993). The melatonin rhythm: both a clock and a calendar. *Experientia* 49, 654–664. doi: 10.1007/bf01923947

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sapède, D., and Cau, E. (2013). The pineal gland from development to function. *Curr. Top. Dev. Biol.* 106, 171–215. doi: 10.1016/B978-0-12-416021-7.00005-5

Sugden, A. L., Sugden, D., and Klein, D. C. (1986). Essential role of calcium influx in the adrenergic regulation of cAMP and cGMP in rat pinealocytes. *J. Biol. Chem.* 261, 11608–11612.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646

Tang, Z. L., Wu, Y., Yang, Y. L., Yang, Y. C. T., Wang, Z. S., Yuan, J. P., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutional conservation in *Sus scrofa*. *Sci. Rep.* 7:43166. doi: 10.1038/srep43166

Theocharidis, A., Van, D. S., Enright, A. J., and Freeman, T. C. (2009). Network visualization and analysis of gene expression data using biolayout express(3D). *Nat. Protoc.* 4, 1535–1550. doi: 10.1038/nprot.2009.177

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Trivedi, P. P., Jena, G. B., Tikoo, K. B., and Kumar, V. (2016). Melatonin modulated autophagy and Nrf2 signaling pathways in mice with colitis-associated colon carcinogenesis. *Mol. Carcinog.* 55, 255–267. doi: 10.1002/mc.22274

UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, 204–212. doi: 10.1093/nar/gku989

Vanlandewijck, M., He, L., Mae, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. *Nature* 554, 475–480. doi: 10.1038/nature25739

Volders, P. J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., et al. (2015). An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.* 43, 174–180.

Wang, J., Garrey, J., and Davis, R. (2014). Transcription in pronuclei and one-to four-cell embryos drives early development in a nematode. *Curr. Biol.* 24, 124–133. doi: 10.1016/j.cub.2013.11.045

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597. doi: 10.1038/nature12364

Yamazaki, F., Møller, M., Fu, C., Clokie, S. J., Zykovich, A., Coon, S. L., et al. (2015). The Lhx9 homeobox gene controls pineal gland development and prevents postnatal hydrocephalus. *Brain Struct. Funct.* 220, 1497–1509. doi: 10.1007/s00429-014-0740-x

Yan, S., Tu, Z., Liu, Z., Fan, N., Yang, H., Yang, S., et al. (2018). A Huntingtin knockin pig model recapitulates features of selective neurodegeneration in huntington's disease. *Cell* 173, 989–1002. doi: 10.1016/j.cell.2018.03.005

Yang, Y., Zhou, R., Zhu, S., Li, X., Li, H., Yu, H., et al. (2017). Systematic identification and molecular characteristics of long noncoding rnas in pig tissues. *Biomed. Res. Int.* 2017:6152582. doi: 10.1155/2017/6152582

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yang, Zhou, Li, Liu, Zhang, Ao, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 June 2019
doi: 10.3389/fgene.2019.00590

Genetic Contribution to Variation in Blood Calcium, Phosphorus, and Alkaline Phosphatase Activity in Pigs

*Henry Reyer[1], Michael Oster[1], Dörte Wittenburg[2], Eduard Murani[1], Siriluck Ponsuksili[3] and Klaus Wimmers[1,4]**

[1] Genomics Unit, Institute for Genome Biology, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, [2] Biomathematics and Bioinformatics Unit, Institute of Genetics and Biometry, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, [3] Functional Genome Analysis Unit, Institute for Genome Biology, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, [4] Department of Animal Breeding and Genetics, Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany

Blood values of calcium (Ca), inorganic phosphorus (IP), and alkaline phosphatase activity (ALP) are valuable indicators for mineral status and bone mineralization. The mineral homeostasis is maintained by absorption, retention, and excretion processes employing a number of known and unknown sensing and regulating factors with implications on immunity. Due to the high inter-individual variation of Ca and P levels in the blood of pigs and to clarify molecular contributions to this variation, the genetics of hematological traits related to the Ca and P balance were investigated in a German Landrace population, integrating both single-locus and multi-locus genome-wide association study (GWAS) approaches. Genomic heritability estimates suggest a moderate genetic contribution to the variation of hematological Ca (N = 456), IP (N = 1049), ALP (N = 439), and the Ca/P ratio (N = 455), with values ranging from 0.27 to 0.54. The genome-wide analysis of markers adds a number of genomic regions to the list of quantitative trait loci, some of which overlap with previous results. Despite the gaps in knowledge of genes involved in Ca and P metabolism, genes like *THBS2*, *SHH*, *PTPRT*, *PTGS1*, and *FRAS1* with reported connections to bone metabolism were derived from the significantly associated genomic regions. Additionally, genomic regions included *TRAFD1* and genes coding for phosphate transporters (*SLC17A1–SLC17A4*), which are linked to Ca and P homeostasis. The study calls for improved functional annotation of the proposed candidate genes to derive features involved in maintaining Ca and P balance. This gene information can be exploited to diagnose and predict characteristics of micronutrient utilization, bone development, and a well-functioning musculoskeletal system in pig husbandry and breeding.

Keywords: minerals, genetics, pigs, genome-wide association, genomic heritability, hematological traits

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Dirk-Jan De Koning, Swedish University of Agricultural Sciences, Sweden
Martin Johnsson, Swedish University of Agricultural Sciences, Sweden

***Correspondence:**
Klaus Wimmers
wimmers@fbn-dummerstorf.de

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *16 November 2018*
Accepted: *04 June 2019*
Published: *28 June 2019*

Citation:
Reyer H, Oster M, Wittenburg D, Murani E, Ponsuksili S and Wimmers K (2019) Genetic Contribution to Variation in Blood Calcium, Phosphorus, and Alkaline Phosphatase Activity in Pigs. Front. Genet. 10:590. doi: 10.3389/fgene.2019.00590

INTRODUCTION

In the body, the homeostasis of calcium (Ca) and phosphorus (P) is maintained to ensure appropriate conditions for bone mineralization, energy utilization, nucleic acid synthesis, and signal transduction of each individual cell and the entire organism. Molecular pathways involved in these processes are regulated by numerous factors such as the parathyroid hormone (PTH), vitamin

D, fibroblast growth factor 23 (FGF23), and the calcium sensing receptor (CASR). The vitamin D system, for example, is able to alter the transcription rates of thousands of target genes *via* vitamin D responsive elements (VDRE) located in their respective promoter region (Pike et al., 2010). In addition, other transcription factors like MafB have been identified as involved in the regulation of mineral homeostasis by orchestrating intracellular signaling (Morito et al., 2018). However, especially with regard to P homeostasis, particular mechanisms of sensing and regulation as well as underlying molecules are still unclear (Chande and Bergwitz, 2018).

In pigs of the same age, the Ca and P levels in the blood differ considerably between different breeds or even within breeds, suggesting a genetic contribution to the variability of mineral concentrations (Rodehutscord, 2001; Hittmeier et al., 2006; Just et al., 2018b). However, reliable heritability estimates for the blood values of Ca and P for pigs are not yet available. An initial understanding of the genetics of the Ca and P homeostasis in pigs was demonstrated by Bovo et al. (2016), who were able to identify quantitative trait loci (QTL) for serum Ca on SSC8, 11, 12, and 13 and for P on SSC2 and 7 in an Italian Large White population. The proposed list of candidate genes emphasizes that the drivers of Ca and P homeostasis are in fact partly unknown and that several factors remain to be identified. Consistently, a recent study on the genetic contribution of well-known functional candidate genes on Ca and P homeostasis in pigs showed only a small contribution of these major players to the genetic variance (Just et al., 2018b). Further insights into the role of genetics in the regulation of Ca and P homeostasis can be derived from human studies on kidney health and bone metabolism (reviewed by Lederer, 2014). Specifically, a GWAS for humans with European ancestry revealed several QTL regions containing functional candidate genes such as *FGF23*, *SLC34A1*, and *CAST*, whereby the most prominent SNPs are located in nearby regions representing other genes (Kestenbaum et al., 2010). Notably, the highest significant association of this particular study was identified for *ALPL*, an alkaline phosphatase that hydrolyzes phosphate compounds at alkaline pH and is involved in bone mineralization (Kestenbaum et al., 2010).

The current GWAS used blood-derived proxies for the Ca and P homeostasis including Ca, inorganic phosphorus (IP), alkaline phosphatase activity (ALP), and the respective Ca/IP ratio. Specifically, ALP represents the total activity of all ALP isoforms in blood and is indicative of endogenous P requirements. The hematological Ca and IP levels as well as the calcium/phosphorus (Ca/P) ratio represent the variation of the strictly regulated mineral balance. In addition, genomic data from pigs were used to estimate genomic heritability and genetic correlations for all analyzed traits. Their availability would be an important prerequisite for assessing the potential of breeding strategies that include (i) the efficient use of Ca and P for bone formation and growth processes, (ii) the prevention of ectopic mineralization of peripheral tissues (hypercalcification), and (iii) the reduction of environmental impacts of livestock farming.

MATERIALS AND METHODS

Pig Population and Phenotypes

Animal care and sampling were carried out in accordance with the guidelines of the German Law of Animal Protection. All protocols have been approved by the Animal Care Committee of the Leibniz Institute for Farm Animal Biology (FBN). Compliance with all relevant international, national, and/or institutional guidelines for the care and use of animals was ensured.

For this study, 1,053 commercial German Landrace pigs have been raised on standard diets (Gesellschaft für Ernährungsphysiologie, 2006) for fattening pigs. The purebred pigs originated from nine different farms in the area of Mecklenburg-Western Pomerania (Germany) and were fattened either at the Institutes pig farm or in the performance test station Jürgenstorf (Germany). Animals had an average age of 163.8 ± 15.5 days (mean ± SD; individual ages ranged from 127 to 222 days). The population consists of 73 males, 355 females, and 625 castrates, which were sired by 64 boars. Pigs had *ad libitum* access to feed and water. Pigs were killed by electrical stunning followed by exsanguination in the experimental slaughterhouse of FBN Dummerstorf. At slaughter, liver samples were collected for DNA extraction. Additionally, trunk blood was sampled for serum and plasma preparation. For the first batch of animals ($N = 590$), IP was measured using heparin plasma. Serum samples were available for the second batch of pigs ($N = 463$) in which levels of Ca, IP, and ALP were measured. Blood chemistry analyses were performed using commercial available assays on a Fuji Dri-Chem 4000i (FujiFilm, Minato, Japan). Phenotypes were transformed to follow a normal distribution by Johnson SU using JMP Genomics 7.0 (SAS Institute, Cary, NC, United States).

Genotyping and Data Preparation

Based on DNA obtained from liver samples, genotyping of the 1053 animals was performed using the 60k porcine SNP bead chip (Illumina, San Diego, CA, United States). Data files were analyzed using the GenomeStudio software (Illumina, version 2.0.3) for clustering of genotypes and initial quality control (sample call rate >95% and SNP call rate >95%). Afterward, missing values in the genotype matrix were imputed with fastPHASE (Scheet and Stephens, 2006). Settings for imputation included 10 runs of the EM algorithm with 50 iterations each and the scanning for genotype errors option was enabled. Genotype errors were excluded by discarding autosomal SNPs with estimated error rate above 10% (Scheet and Stephens, 2006). SNP sequences of the bead chip were mapped to pig genome assembly 11.1 (accessed on July 13, 2017) using Bowtie2 (version 2.2.6). Markers not mapping to autosomes in the current genome assembly were dropped. Additional filtering of markers was applied at the level of Hardy–Weinberg equilibrium ($P > 1 \times 10^{-6}$) and, after excluding individuals with missing phenotypes for the corresponding trait, for minor allele frequency (MAF <0.03). In total, the number of markers was reduced from 61,565 to 47,946 (IP; 1049 pigs), 47,302 (Ca; 456 pigs), 47,298 (Ca/P; 455 pigs), and 47,258 (ALP; 439 pigs), respectively.

Estimation of Genetic and Phenotypic Parameters

Bayesian estimates of genomic heritability were obtained with the help of trait-specific univariate models using the R package BGLR version 1.0.5 (Pérez and de Los Campos, 2014). The polygenic effect was included in the model employing the genomic relationship matrix (VanRaden, 2008). Furthermore, models included age of pigs as covariate and, additionally, batch effect for IP. Genetic correlations between traits were analyzed in a bivariate model using the MTM package version 1.0.0[1]. Due to the different total number of animals available for a particular trait, the smallest common overlap in the marker set was used in each trait combination. For both estimation of genomic heritability and genetic correlation, analysis parameters have been set to default values. The output from 200,000 iteration steps was used. Diagnosis of convergence was done using the Gelman–Rubin function applied to additional Markov chains; it is implemented in the coda R package version 0.19-1. The potential scale reduction factor was <1.1, indicating convergence (Gelman and Rubin, 1992). After the first 50,000 iterations were discarded as burn-in phase, the average genomic heritability ± SD and correlation coefficient ± SD were estimated from the remaining iterations. Moreover, phenotypic correlations, expressed as Pearson coefficients, were calculated for the traits transformed to follow a normal distribution as described above.

Genome-Wide Association Study

Both single-locus and multi-locus genome-wide association study (GWAS) approaches were used in this study to identify genomic regions contributing to the genetic variance of blood Ca, IP, and ALP levels and the Ca/P ratio. The single-locus GWAS was performed using a mixed linear model implemented in JMP Genomics 7.0 (SAS Institute). For all traits, the sire was included as a random effect in the model, which not only accounts for relatedness in the population but also controls for environmental factors (farm effect). Additionally, the age of pigs was used as a covariate in the model. The analysis of IP further included batch as fixed effect. Due to the linkage disequilibrium between markers, the analysis of a single locus was not necessarily independent from other loci, which was taken into account by using SimpleM to estimate the number of independent tests (Gao et al., 2008). SimpleM revealed a number of 19,773 independent tests for the entire dataset. Accordingly, significance thresholds were set at 1/19,773 [$-\log_{10}$ (P-value) = 4.30] for suggestive significance and 0.05/19,773 [$-\log_{10}$ (P-value) = 5.60] for genome-wide significance (Lander and Kruglyak, 1995). To investigate model fit, a quantile–quantile (QQ) plot based on the observed P-values was plotted for each trait. Manhattan plot representation of genome-wide results was created with the postGWAS R package (version 1.11-2) (Hiersche et al., 2013). Data for specific genomic regions were visualized using LocusZoom (version 1.3) (Pruim et al., 2010).

[1]https://github.com/QuantGen/MTM

The multi-locus GWAS was based on a Bayesian variable selection approach (Bayes B; Meuwissen et al., 2001) implemented in the BGLR package (Pérez and de Los Campos, 2014). In agreement with single-locus GWAS, models included age of pigs as covariate for all traits and batch effect for the analysis of IP and sire. A total of 200,000 cycles of the Gibbs sampler were performed after convergence diagnosis (as described above), with the first 25% of the iterations discarded as a burn-in phase. For the Bayes B algorithm, the prior proportion of non-zero marker effects was set to 0.5% for the analysis of Ca and Ca/P, as previously described (Reyer et al., 2015; Fernando et al., 2017). Based on genomic heritability estimates, which were considerably higher for IP and ALP, the prior proportion of non-zero marker effects was increased to 1% for these traits. In total, 468.2 (IP), 199.0 (Ca), 192.0 (Ca/P), and 433.0 (ALP) markers were considered on average in each cycle of the Gibbs sampler. Other parameters were set to default values. Bayes factors (BFs) were calculated based on the posterior probability of inclusion for each marker and expressed as the quotient between posterior and prior odds ratio (Karkkainen and Sillanpaa, 2012). Markers with a BF > 10 were considered as having decisive evidence according to Jeffreys (1998).

Data Integration and Candidate Genes

Based on the combination of results from single- and multi-locus GWAS, genes were explored in the proximity of significant markers (closest genes in each direction plus their overlapping genes) using the R package postGWAS. To select genes linked to SNPs, linkage disequilibrium (LD) between markers was calculated based on the genotype matrix using the snp2gene function (Hiersche et al., 2013). Those genes were revealed, which were in LD with SNPs significantly associated with a trait. Specifically, genes were considered where the maximum LD between one of the gene-representing SNP and the significantly associated SNP was above 0.6. A window of 1 Mb around the significantly associated SNP was examined. In addition, the 95% confidence interval of a QTL region was estimated using the likelihood approach proposed by Li (2011). The names of genes that belong to these intervals were extracted. Subsequently, all resulting gene names (from proximity and LD analyses and CI intervals) were converted into human orthologous gene identifiers using the biomaRt R package (version 2.34.2). Gene lists were combined for all traits and passed to the ClueGO (v2.5.1) Cytoscape (v3.6.1) plugin for analysis of gene ontology (GO) (Bindea et al., 2009). The following databases were included: GO cellular component, GO molecular function, and REACTOME Pathways (all accessed on June 6, 2018). ClueGO settings were as follows: GO tree interval of 2–8, cutoff of more than or equal to 4 genes and 3% associated genes, and kappa score (κ) = 0.4. GO term enrichment was tested with a two-sided hypergeometric test considering a Benjamini–Hochberg adjusted P-value ≤0.05. In addition, the QTL regions revealed from both GWAS approaches were manually screened for positional (in proximity to highest significantly associated SNPs) and functional (known function related to the trait of

interest) candidate genes. Therefore, genomic information and functional annotation of genes were retrieved from Ensembl[2] and GeneCards – the human gene database[3], respectively (Stelzer et al., 2016).

RESULTS

All traits showed a considerable variation in the examined population of German Landrace pigs (**Table 1**). Phenotypic correlations between levels of Ca, IP, and ALP were all positive and showed a low to moderate magnitude (**Table 2**). A high negative phenotypic correlation coefficient was found between IP and Ca/P, while the correlation between Ca and Ca/P was moderate and positive. Almost no correlation was revealed between ALP and Ca/P. Regarding the genetic correlation, values were positive for Ca-IP, IP-ALP, and Ca-Ca/P, whereas correlations for ALP-Ca/P and ALP-Ca were negative (**Table 2**). The corresponding correlation coefficients showed a low to moderate magnitude. However, high standard deviations of these estimations indicate a high degree of uncertainty. According to the phenotypic level, the highest negative genetic correlation was observed between IP and Ca/P at −0.62. In general, all traits showed a moderate genomic heritability (**Table 2**). The highest genomic heritability estimates were obtained for ALP (h^2 = 0.54) and IP (h^2 = 0.42). Estimates for Ca and Ca/P were considerably lower at 0.27.

The most promising genomic regions associated with IP were identified on SSC6 and 14 (**Table 3**). Specifically, the QTL on SSC6 at 110 Mb was commonly indicated by the two GWAS approaches. Genes closest to the highest significantly associated marker, ASGA0090429 (*P*-value = 5.24, BF = 14.22), are GATA Binding Protein 6 (*GATA6*) and RB Binding Protein 8 (*RBBP8*) (**Figure 1**). ALGA0036329, obtained by multi-locus analysis, showed the highest BF (BF = 27.3) and was also located in this QTL region on SSC6. Single-locus analysis revealed ALGA0077099 (rs80988848) as the highest significantly associated SNP for IP (*P*-value = 6.24). It mapped to BICD Family Like Cargo Adaptor 1 (*BICDL1*) at 40.12 Mb on SSC14, whereas the indicated genomic region harbors several putative candidate genes (**Supplementary Figure 1**). Additionally, the QTL region located on SSC1 between 109.6 and 111.2 Mb was indicated by five significantly associated SNPs, whereas MARC0015485 reached genome-wide significance (*P*-value = 5.89). The SNP is located between Vacuolar Protein Sorting 13 Homolog C (*VPS13C*) and RAR Related Orphan Receptor A (*RORA*).

For Ca, genomic regions were identified on SSC3, 6, 11, 13, 16, and 18 (**Table 4**). Exclusively, the QTL on SSC6 at 7 Mb was identified by the two GWAS approaches, with the highest significantly associated SNP, ALGA0104738 (*P*-value = 4.52, BF = 10.48) located between Gigaxonin (*GAN*) and Beta-Carotene Oxygenase 1 (*BCO*). Two additional regions on SSC6, at 10.8 and 11.1 Mb, were indicated by multi-locus analysis bordered by *ENSSSCG00000039747* (metalloproteinase inhibitor 1-like) and Contactin Associated Protein Like 4 (*CNTNAP4*). Three SNPs, which reached the suggestive significant level for association with Ca, were highlighted on SSC13 and indicate Family With Sequence Similarity 43 Member A (*FAM43A*) and Xyloside Xylosyltransferase 1 (*XXYLT1*) as positional candidate genes.

The highest significant association in single-locus analysis for Ca/P was identified on SSC8 (**Table 5** and **Figure 1**). The corresponding SNP ALGA0118376 mapped in intron 6 of Leucine Rich Repeat LGI Family Member 2 (*LGI2*). Moreover, several significantly associated SNPs were identified on SSC14 at 33 Mb and between 84 and 87 Mb. Respective positional candidate genes are *ENSSSCG00000010344* and Glutamate Ionotropic Receptor Delta Type Subunit 1 (*GRID1*) located around 84.3 and 86.9 Mb. On SSC13, ASGA0056810 showed highest contribution to the genetic variance in multi-locus analysis (BF = 22.4) and mapped in an intronic region of Unc-51 Like Kinase 4 (*ULK4*). The QTL on SSC17 is indicated by two significantly associated SNP (*P*-value >4.3, BF > 3) pointing to Protein Tyrosine Phosphatase, Receptor Type T (*PTPRT*) as candidate.

SNPs with genome-wide significance in single-locus analysis for ALP were identified on SSC7 at 17.4 and 21.4 Mb (**Table 6**). Additionally, the region harboring the latter SNP was also indicated by multi-locus analysis. Putative candidate genes located nearby the lead SNPs were SRY-Box 4 (*SOX4*), prolactin (*PRL*), POM121 Transmembrane Nucleoporin Like 2 (*POM121L2*), and Zinc Finger Protein 184 (*ZNF184*), while the highest significantly associated SNP ALGA0039405 (*P*-value = 6.75, BF = 17.74) mapped in an intronic region of the Serine Protease 16 gene (*PRSS16*) (**Supplementary Figure 2**). Interestingly, in between these QTL at approximately 20.4–20.7 Mb, a cluster of P transporters (*SLC17A1–SLC17A3*) is located. Further genomic regions with putative effect on ALP were identified at 11.8 and 24.2 Mb also on SSC7. Furthermore, QTL commonly identified by the two GWAS approaches were

[2]http://www.ensembl.org

[3]www.genecards.org

TABLE 1 | Descriptive statistics of blood traits in German Landrace pigs.

Trait	Acronym	*N*	Mean	SD	Min	Max
Inorganic phosphorus (mg/dl)	IP	1049	8.72	1.11	6.0	12.7
Calcium (mg/dl)	Ca	456	10.00	0.78	6.8	11.9
Calcium/phosphorus ratio	Ca/P	455	1.10	0.13	0.81	1.73
Alkaline phosphatase (U/L)	ALP	439	118.95	31.93	32	250

TABLE 2 | Estimates of genetic (below the diagonal) and phenotypic (above the diagonal) correlation coefficients and the genomic heritability (diagonal, bold) for concentrations of IP, Ca, ALP, and Ca/P ratio.

Trait	IP	Ca	ALP	Ca/P
IP	**0.42 ± 0.05**	0.26	0.14	−0.75
Ca	0.23 ± 0.22	**0.27 ± 0.07**	0.11	0.40
ALP	0.06 ± 0.20	−0.16 ± 0.21	**0.54 ± 0.08**	−0.03
Ca/P	−0.62 ± 0.13	0.29 ± 0.20	−0.21 ± 0.20	**0.27 ± 0.07**

TABLE 3 | Genomic regions and corresponding lead SNPs identified by single- and multi-locus GWAS for levels of IP.

Lead SNP	SSC[1]	SNP position	95% confidence interval (Mb)[2]	$-\log_{10}$ (*P*-value)	Bayes factor	MAF[3]	% Var[4]
M1GA0000151	1	887856	0.888 ± 0.003	4.44	4.27	0.246	1.6
ASGA0106092	1	109579245	109.579 ± 0.002	5.04	5.26	0.272	1.9
MARC0015485	1	110706305	110.707 ± 0.002	5.89	5.30	0.130	2.2
ASGA0103590	2	41032321	41.033 ± 0.001	4.38	8.96	0.414	1.6
ASGA0083196	3	21735144	21.734 ± 0.005	3.86	16.43	0.248	1.4
ASGA0026923	5	96278945	96.280 ± 0.002	3.28	12.22	0.265	1.2
ALGA0036329	6	109768450	109.768 ± 0.001	4.82	27.32	0.435	1.8
ASGA0090429	6	109873081	109.875 ± 0.005	5.24	14.22	0.491	2.0
ASGA0094586	6	114804166	114.802 ± 0.004	4.45	1.96	0.341	1.6
ALGA0054999	9	121587586	121.589 ± 0.003	4.39	3.65	0.111	1.6
SIRI0000950	14	11099200	11.100 ± 0.000	4.74	6.19	0.352	1.7
ALGA0077099	14	40117526	40.104 ± 0.026	6.24	4.61	0.489	2.4
M1GA0018562	14	40465628	40.807 ± 0.342	5.58	3.03	0.461	2.1
ALGA0088210	15	134132243	134.133 ± 0.001	4.89	7.26	0.406	1.8

[1]Sus scrofa chromosome; [2]Calculated according to Li's method (Li, 2011); [3]Minor allele frequency; [4]Explained proportion of the phenotypic variance by the lead SNP.

located on SSC1 (262.4 Mb), SSC5 (74.0 Mb), SSC8 (74.2 Mb), and SSC9 (64.9 Mb). Corresponding positional candidate genes for these QTL are shown in **Figure 1**. QQ plots for single-locus analyses of all traits are shown in **Supplementary Figure 3**, which indicate an inflation of *P*-values. While a similar inflation was observed using a polygenic effect combined with a genomic relationship matrix instead of a sire effect (results not shown), the causal factors for this inflation remain unknown.

For data integration and analysis of GO, a list of genes was derived from single- and multi-locus GWAS. Therefore, genes located in the 95% confidence intervals of QTL regions and genes that are in LD with significantly associated SNPs were considered. A list of 191 candidate genes was obtained and used for GO term analysis. In total, 61 GO terms reached statistical significance (adjusted *P*-value $\leq$0.05) and clustered in 7 GO groups (**Figure 2**). The largest group was mainly formed by genes of the histone cluster family, which resulted in enriched GO terms such as "chromatin organization," "nucleosome," and "protein–DNA complex." Genes located in QTL regions for traits related to Ca–P balance were further enriched for GO terms such as "cell–cell communication," "phosphatidylinositol bisphosphate binding," and "smooth muscle contraction." Interestingly, due to the cluster of solute carriers located in the QTL region on SSC7, "solute:sodium symporter activity" was also enriched.

DISCUSSION

Genomic Heritability of Phenotypes Related to Ca and P Homeostasis

In human diagnostics, hematological parameters can be determined easily and are a valuable tool for assessing the patient's state of health. Parameters related to the Ca and P balance are used to indicate mineral status and bone turnover and to contribute to the diagnosis of bone health, vascular calcification, and kidney diseases (Kestenbaum et al., 2010). Corresponding phenotypes are derived from large cohorts and allow comprehensive studies to elucidate the genetic contribution to the variation in these traits. In terms of heritability, the estimates for serum Ca (0.33) and IP (0.58) obtained from these human studies correspond well in magnitude to the genetic contribution investigated in this study for pigs (Hunter et al., 2002). For IP, the estimate of heritability was almost twice as high as for serum Ca in both species, which might indicate conserved mechanisms to maintain Ca and P homeostasis within narrow ranges. The serum ALP activity, which showed the highest genomic heritability among the traits analyzed in this study, was also attributed a high heritability in humans (Nielson et al., 2012). Moreover, breed-specific differences in ALP were mentioned for cattle (Cole et al., 2001). The total activity of ALP in serum is indicative for concentrations of local ALP isoforms from liver, muscle, bone, and bile duct. Its serum values point to abnormalities of the corresponding tissues; e.g., with regard to bones, it serves as early marker for increased bone turnover (Hoffmann and Solter, 2008). The genomic heritability estimate for the Ca/P ratio of the analyzed pigs was moderate and in similar magnitude to genomic heritability for Ca. Ca/P is proposed as an indicator of bone mobilization and reflects the P status (Anderson et al., 2017). Indeed, Ca/P showed high negative correlation with IP and moderate positive correlation with Ca at the phenotypic and genotypic level. This is very much in accordance with correlations between IP and the serum Ca × P product in humans (Yokoyama, 2008). Ca and IP showed a moderate positive phenotypic and genetic correlation in this study, albeit the estimated genetic correlation is accompanied by some uncertainty as indicated by the high standard deviations of these estimations. The positive correlations reflect the organism's efforts to maintain both Ca and IP levels, as well as their ratio, at a certain level (Feher, 2017). Moreover, the common genetics of both traits, indicated by the moderate genetic correlation, is, to some extent, expected considering the complex interplay between both minerals and the numerous factors that affect their

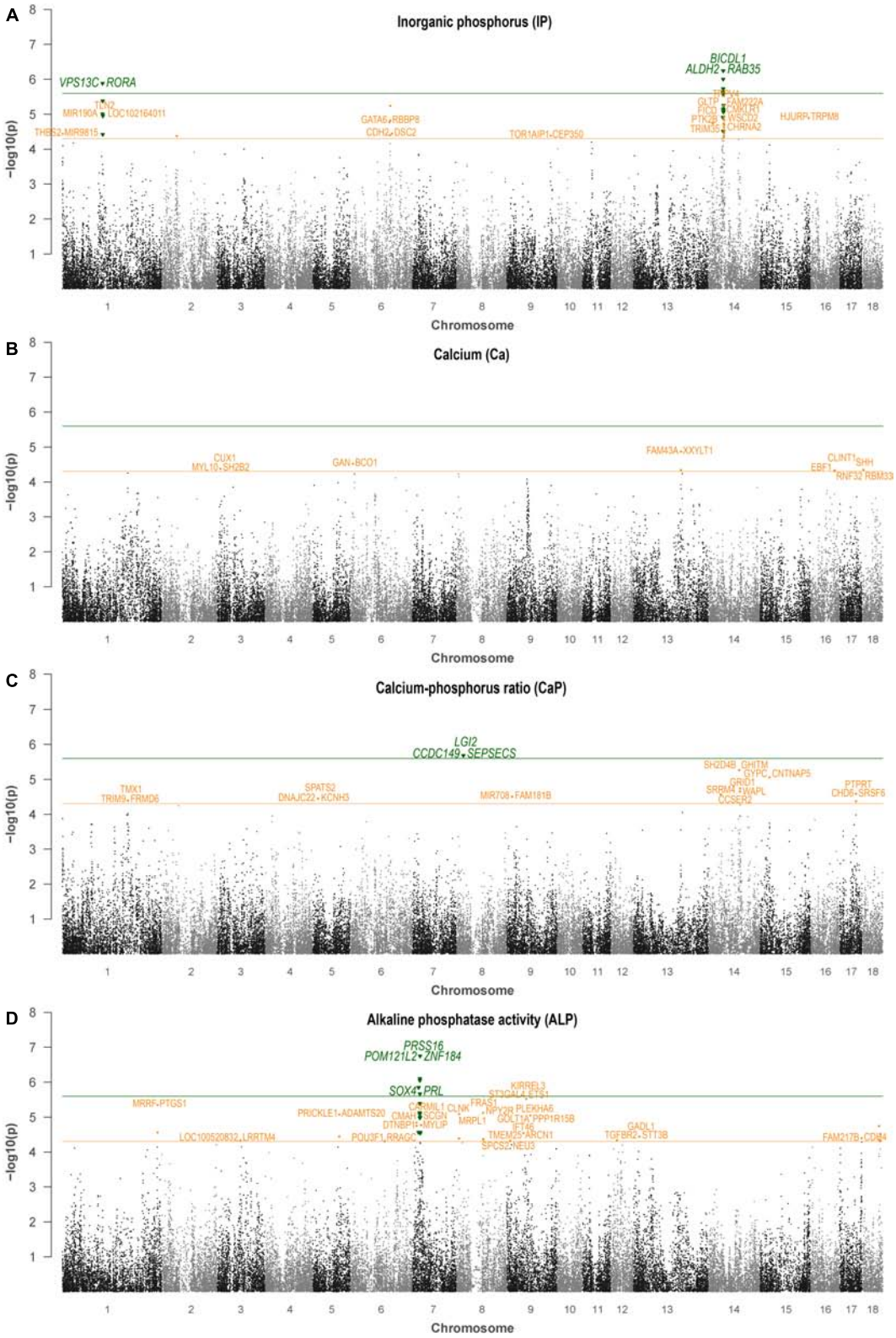


FIGURE 1 | Single-locus association analysis of hematological levels of IP **(A)**, calcium **(B)**, calcium/phosphorus ratio **(C)**, and ALP **(D)**. Orange and green lines indicate suggestive significance and genome-wide significance, respectively. Genes located closest to the lead SNP of a 2-Mb genomic windows are indicated.

TABLE 4 | Genomic regions and corresponding lead SNPs identified by single- and multi-locus GWAS for serum levels of Ca.

Lead SNP	SSC[1]	SNP position	95% confidence interval (Mb)[2]	$-\log_{10}$ (*P*-value)	Bayes factor	MAF[3]	% Var[4]
MARC0088651	3	9431382	9.432 ± 0.001	4.38	5.82	0.074	3.6
ALGA0104738	6	6998594	7.000 ± 0.001	4.52	10.48	0.249	3.8
ASGA0094980	6	10776875	10.777 ± 0.001	3.99	12.97	0.107	3.3
ASGA0088451	6	11145506	11.188 ± 0.043	4.24	11.12	0.059	3.5
ALGA0107249	11	38134199	38.134 ± 0.001	3.65	11.17	0.209	3.0
MARC0004520	13	131771363	131.801 ± 0.030	4.87	6.63	0.242	4.1
ALGA0091318	16	65219767	65.219 ± 0.001	4.33	5.76	0.347	3.6
ASGA0078611	18	2557956	2.558 ± 0.000	4.35	1.20	0.102	3.6

[1]*Sus scrofa chromosome;* [2]*Calculated according to Li's method (Li, 2011);* [3]*Minor allele frequency;* [4]*Explained proportion of the phenotypic variance by the lead SNP.*

TABLE 5 | Genomic regions and corresponding lead SNPs identified by single- and multi-locus GWAS for the Ca/P ratio.

Lead SNP	SSC[1]	SNP position	95% confidence interval (Mb)[2]	$-\log_{10}$ (*P*-value)	Bayes factor	MAF[3]	% Var[4]
ASGA0000047	1	477400	0.477 ± 0.002	3.88	10.59	0.424	3.2
M1GA0001253	1	180934106	180.936 ± 0.005	4.39	2.66	0.477	3.7
ALGA0030823	5	15372183	15.372 ± 0.001	4.45	11.77	0.320	3.7
ALGA0118376	8	18998044	18.999 ± 0.000	5.68	1.84	0.054	4.9
ASGA0041852	9	14569966	14.570 ± 0.000	4.50	1.37	0.277	3.8
ASGA0056810	13	25360591	25.361 ± 0.001	3.54	22.43	0.490	2.9
MARC0060323	14	33722981	33.724 ± 0.000	4.57	5.00	0.398	3.8
INRA0045347	14	84495128	84.496 ± 0.002	5.26	8.31	0.314	4.5
INRA0045383	14	86273463	86.898 ± 0.003	4.65	5.24	0.255	3.9
ALGA0113690	15	26316801	26.316 ± 0.002	5.05	7.25	0.281	4.3
MARC0009678	17	44751346	44.753 ± 0.002	4.59	4.43	0.163	3.8

[1]*Sus scrofa chromosome;* [2]*Calculated according to Li's method (Li, 2011);* [3]*Minor allele frequency;* [4]*Explained proportion of the phenotypic variance by the lead SNP.*

TABLE 6 | Genomic regions and corresponding lead SNPs identified by single- and multi-locus GWAS for serum ALP.

Lead SNP	SSC[1]	SNP position	95% confidence interval (Mb)[2]	$-\log_{10}$ (*P*-value)	Bayes factor	MAF[3]	% Var[4]
ALGA0010101	1	262433517	262.434 ± 0.001	5.35	14.44	0.481	4.7
H3GA0004746		262469788		4.56	24.83	0.417	4.0
MARC0075394	3	65151368	65.152 ± 0.002	4.33	1.92	0.372	3.7
MARC0008087	5	73977137	73.980 ± 0.002	5.08	18.54	0.236	4.5
H3GA0018478	6	94536445	94.534 ± 0.004	4.32	5.42	0.165	3.7
ALGA0038660	7	11794308	11.795 ± 0.001	4.84	4.73	0.244	4.2
H3GA0020150	7	17440108	17.441 ± 0.003	5.85	2.31	0.087	5.2
ALGA0039405	7	21385114	21.376 ± 0.010	6.75	17.74	0.192	6.1
ASGA0032054	7	24178503	24.179 ± 0.000	4.77	2.92	0.491	4.2
ALGA0109925	8	7328296	7.329 ± 0.001	5.09	6.15	0.418	4.5
ASGA0039012	8	73863861	73.864 ± 0.002	5.12	4.45	0.437	4.5
ALGA0102419	8	74229267	74.230 ± 0.003	4.37	11.76	0.427	3.8
ASGA0090661	9	9326958	9.319 ± 0.008	4.32	2.22	0.245	3.7
DIAS0002588	9	45853528	45.854 ± 0.001	4.54	3.36	0.068	3.9
H3GA0027430	9	53561115	53.561 ± 0.001	5.52	6.27	0.462	4.9
ASGA0043611	9	64902900	64.920 ± 0.017	5.05	14.42	0.141	4.4
ASGA0101422	12	16899513	16.901 ± 0.004	4.34	2.40	0.134	3.7
SIRI0000547	13	17031957	17.032 ± 0.002	4.44	1.37	0.166	3.8
MARC0038389	16	10488789	10.490 ± 0.002	3.53	14.95	0.090	3.0
MARC0009297	17	60317999	60.315 ± 0.003	4.39	3.32	0.349	3.8
MARC0055314	18	45755314	45.756 ± 0.002	4.75	6.62	0.227	4.1
ALGA0119274	18	48412007	48.410 ± 0.003	4.51	2.66	0.030	3.9

[1]*Sus scrofa chromosome;* [2]*Calculated according to Li's method (Li, 2011);* [3]*Minor allele frequency;* [4]*Explained proportion of the phenotypic variance by the lead SNP.*

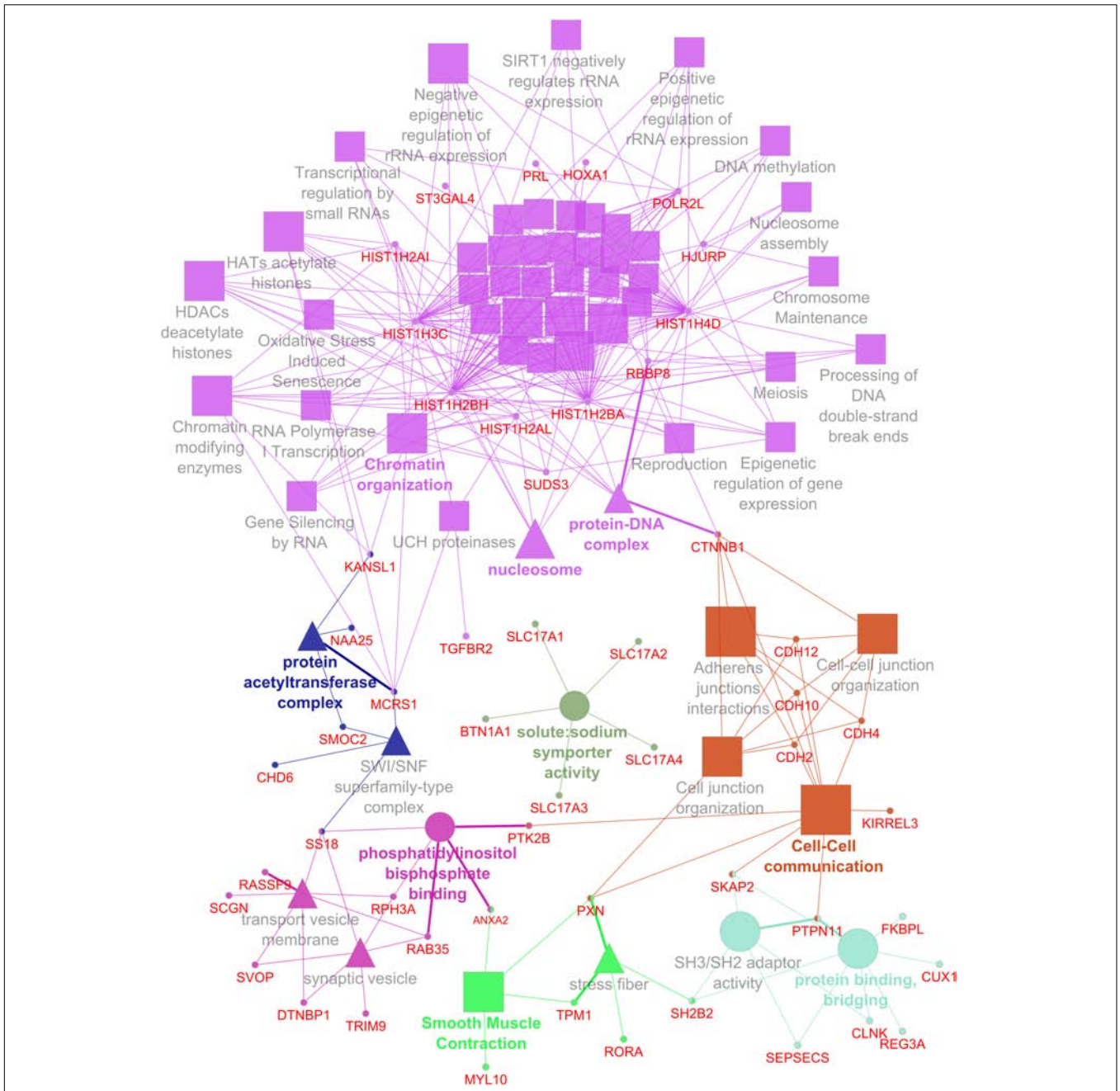


FIGURE 2 | Network representing enriched GO terms and linked genes based on QTL for hematological traits related to the calcium and phosphorus homeostasis.

regulation such as PTH, FGF23, and vitamin D metabolites (Shaker and Deftos, 2018).

Genomic Regions Associated With Hematological Traits Related to the Ca and P Balance

Existing genome-wide analyses in different species indicate mainly positional candidate genes with not yet known function in Ca and P homeostasis (Reiner et al., 2009; Kestenbaum et al., 2010; Bovo et al., 2016; Van Goor et al., 2016). Notably, taking into account the QTL intervals derived from a human meta-analysis, at least some functional candidate genes such as *CASR*, *FGF23*, *ALPL*, and *SLC34A1* are included (Kestenbaum et al., 2010). In pigs, even the targeted association analysis of obvious candidate genes involved in Ca and P regulation could not demonstrate significant contributions to the phenotypic variability (Just et al., 2018b). Similarly, in the current GWAS, none of the leading SNPs of the QTL point to positional candidate genes that are hitherto known to be major players

in the regulation of Ca and P homeostasis. Nevertheless, the current study revealed several QTL regions for the analyzed traits, which partly overlap with genomic regions from pig QTL database. Specifically, porcine genome regions with a contribution to the traits analyzed have been previously mapped on different chromosomes (**Supplementary Table 1**; Reiner et al., 2009; Yoo et al., 2012; Bovo et al., 2016; Just et al., 2018b). Interestingly, despite the physiological relationship within mineral homeostasis, only few overlapping genomic regions between the individual traits have been identified.

Candidate Genes Associated With IP Levels

Promising candidate genes for IP, due to their proximity and linkage to leading SNPs, are *BICDL1* and Ras-related protein Rab-35 (*RAB35*) on SSC14 (**Supplementary Figure 1**). According to current knowledge, both *BICDL1* and *RAB35* are involved in the regulation of neurite outgrowth. Indeed, processes involved in the extension of neurons are dependent on Ca entries and phosphorylation events (Sutherland et al., 2014). However, whether changes in these processes are detectable at blood level is questionable. In addition, *RAB35* is a key regulator of intracellular membrane transport and involved in endocytosis. The same QTL on SSC14 further contains the recently proposed candidate gene TRAF-Type Zinc Finger Domain Containing 1 (*TRAFD1*) (Just et al., 2018b), but also harbors several other genes demanding the dissection of this QTL to provide further insights. Although the genomic region on SSC6 was the most prominent in the multi-marker analysis for IP, the positional candidate genes (*GATA6* and *RBBP8*) lack any connection to P homeostasis so far. The candidate gene on SSC1, *THBS2*, is proposed due to its functions in mediating cell-to-cell interaction and inhibiting angiogenesis, even though the association only reached the suggestive significance level. With respect to Ca and P homeostasis, mice lacking *THBS2* showed altered bone growth including increased bone density and cortical thickness (Kyriakides et al., 1998).

Candidate Genes Associated With Ca Levels

For Ca, which showed the lowest genomic heritability among the traits analyzed, only few significantly associated genomic regions were detected by the two GWAS methods. Interestingly, genomic regions on SSC6 and SSC18 border QTL previously identified by microsatellite analysis (Reiner et al., 2009). Based on the positional overlap or proximity to leading SNPs, Cut Like Homeobox 1 (*Cux1*; SSC3), *GAN* (SSC6), *BCO1* (SSC6), and Sonic Hedgehog (*SHH*; SSC18) were proposed as positional candidate genes. Considering the currently known functional involvement of these genes, *BCO1* and *SHH* in particular have relations to bone metabolism and health and might be screened for genetic variations. Specifically, *BCO1* represents a key enzyme in the metabolism of vitamin A that also affects bone formation and calcium metabolism (Frankel et al., 1986; Binkley and Krueger, 2000). SHH is associated to the initiation of osteogenesis through interactions with bone morphogenetic proteins (BMP) (Yuasa et al., 2002).

Candidate Genes Associated With the Ca/P Ratio

The highest significant association resulting from single- and multi-locus analysis of Ca/P pointed to SSC8 and SSC13 with *LGI2* and *ULK4* as positional candidate genes. Both genes are reported to be widely expressed in different tissues (see text footnote 3). So far, little is known about the function of LGI2, apart from its association with epilepsies (Limviphuvadh et al., 2010). ULK4 has several functions in the brain including involvement in neuronal cell proliferation and cell-cycle regulation, whereas functions outside the central nervous system are largely unknown (Liu et al., 2017). Considering current functional information, the GWAS results propose Thioredoxin Related Transmembrane Protein 1 (*TMX1*) on SSC1 and Receptor-Type Tyrosine-Protein Phosphatase T (*PTPRT*) on SSC17 as the most interesting positional and functional candidate genes. *TMX1* is highlighted for its role in the regulation of Ca pumps at the contact surface between mitochondria and endoplasmic reticulum, thus influencing Ca transfer and mitochondria activity (Krols et al., 2016). Tyrosine-Protein Phosphatases are known to play a central role in the formation of bone, specifically in processes such as osteoclast production and function and RANKL-mediated signaling (Hendriks et al., 2013). So far, *PTPRT* was shown to cause obesity with altered insulin resistance and lowered feed intake in a knock-out mouse model, whereas in cattle, it was associated with meat quality traits (Tizioto et al., 2013; Feng et al., 2014).

Candidate Genes Associated With ALP Activity

Several genomic regions identified for ALP were located in or near the QTL recorded in the QTL database. Specifically, regions on SSC6 and SSC7 overlap with two of the main findings of the study by Reiner et al. (2009). For the QTL on SSC6, *ALPL* at 79.6 Mb was initially proposed as positional candidate and highlighted for its functional role in bone mineralization in humans (Nielson et al., 2012). The QTL identified in the current study, however, pointed to a region around 94.5 Mb where no known gene mapped. The region on SSC7 harbors the highest significantly associated SNP obtained from single-locus analyses. ALGA0039405 is located in *PRSS16*, which acts in T-cell development and antigen-presenting pathways and is associated with human diabetes susceptibility (Guerder et al., 2018). Taking into account genes containing at least suggestive significant markers, the list of positional candidates in this QTL region further comprises phosphate transporters (*SLC17A1* and *SLC17A4*) and nucleosome-related genes (*HIST1H3E*, *HIST1H1D*, and *HMGN4*). In particular, the genetics of phosphate transporters are worth analyzing, as these play an important role in the Ca and P balance, even though corresponding polymorphisms have so far only been associated with gout and cholesterol homeostasis (Dehghan et al., 2008; Koyama et al., 2015). Other positional candidate genes mapped in the QTL identified for SSC1 and SSC8. H3GA0004746 on SSC1 revealed the highest BF for ALP and is an intron variant of the Prostaglandin-Endoperoxide Synthase 1 (*PTGS1*). PTGS1 is involved in prostaglandin metabolism and angiogenesis.

Moreover, prostaglandins are known to affect bone metabolism (Blackwell et al., 2010). *PTGS1* was recently proposed as a candidate gene for ankylosing spondylitis, a disease accompanied by bone overgrowth (Cortes et al., 2015). The QTL on SSC8 was indicated by both single- and multi-locus GWAS and pointed to Fraser Extracellular Matrix Complex Subunit 1 (*FRAS1*) as positional candidate. Observations that patients with *FRAS1* mutations could have more frequent skull ossification defects (van Haelst et al., 2008) tie in with evidence of an involvement of this gene in familial sclerosing bone dysplasia revealed by exome sequencing (Gannagé-Yared et al., 2014).

Molecular Mechanisms Contributing to the Ca and P Homeostasis

The final interpretation of the gene list derived from the current GWAS suffers due to considerable gaps in the functional annotation of proposed candidate genes. Although most putative positional candidates showed indications of involvement in the Ca and P balance and bone metabolism according to literature, many of these genes are not yet assigned to corresponding GO terms and thus not fully considered. Nevertheless, GO term analysis revealed some obvious terms in connection with the Ca and P balance. Phosphatidylinositol pathways, for example, have previously been described as affected in the context of altered dietary Ca and P intake in pigs (Just et al., 2018a). Similarly, P transporters and the role of Ca in the contraction of smooth muscles are molecular themes that can make a significant contribution to the genetic variance of the traits analyzed (Jiang and Stephens, 1994; Kestenbaum et al., 2010). The large proportion of GO terms related to nucleosomes is mainly driven by the porcine histone gene cluster on SSC7 and reflects the already improved functional annotation available for these genes. The other GO terms emphasized genes that are involved in cellular signaling, cell communication, and posttranslational modification and thus mainly represent intracellular actions of Ca and P. It should be noted that the extracellular Ca concentration is around 20,000 times higher compared to intracellular levels (Clapham, 2007). However, there is evidence that the sensing of intracellular levels might trigger pathways that also affect the extracellular Ca concentration (Bronner, 2001; Just et al., 2018a).

CONCLUSION

The current study elucidates the genetic parameters of Ca, IP, Ca/P, and ALP and provides a list of positional and functional candidate genes and QTL regions for further dissection. The consideration of the results might prove beneficial in relation to pig breeding for both a more efficient utilization of dietary minerals and for an optimal development and maintenance of the skeletal system.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the German Law of Animal Protection. All protocols have been approved by the Animal Care Committee of the Leibniz Institute for Farm Animal Biology (FBN).

AUTHOR CONTRIBUTIONS

KW and EM designed and supervised the study. EM and SP collected the data. HR and MO conducted the experiments. HR, DW, and MO analyzed the data. HR wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This study has received funding from the European Research Area Network on Sustainable Animal Production (ERA-NET SusAn) as part of the Pegasus Project (2817ERA02D). The Leibniz Institute for Farm Animal Biology (FBN) has provided its own funding. The publication of this article was funded by the Open Access Fund of the Leibniz Institute for Farm Animal Biology (FBN).

ACKNOWLEDGMENTS

We thank Angela Garve, Hannelore Tychsen, and Nicole Gentz for their excellent technical help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00590/full#supplementary-material

REFERENCES

Anderson, S., Kidd, L., Benvenutti, M., Fletcher, M., and Dixon, R. (2017). New candidate markers of phosphorus status in beef breeder cows. *Anim. Prod. Sci.* 57, 2291–2303. doi: 10.1071/AN17363

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi: 10.1093/bioinformatics/btp101

Binkley, N., and Krueger, D. (2000). Hypervitaminosis A and bone. *Nutr. Rev.* 58, 138–144. doi: 10.1111/j.1753-4887.2000.tb01848.x

Blackwell, K. A., Raisz, L. G., and Pilbeam, C. C. (2010). Prostaglandins in bone: bad cop, good cop? *Trends Endocrinol. Metab.* 21, 294–301. doi: 10.1016/j.tem.2009.12.004

Bovo, S., Schiavo, G., Mazzoni, G., Dall'olio, S., Galimberti, G., Calò, D., et al. (2016). Genome-wide association study for the level of serum electrolytes in Italian large white pigs. *Anim. Genet.* 47, 597–602. doi: 10.1111/age.12459

Bronner, F. (2001). Extracellular and intracellular regulation of calcium homeostasis. *Sci. World J.* 1, 919–925. doi: 10.1100/tsw.2001.489

Chande, S., and Bergwitz, C. (2018). Role of phosphate sensing in bone and mineral metabolism. *Nat. Rev. Endocrinol.* 14, 637–655. doi: 10.1038/s41574-018-0076-3

Clapham, D. E. (2007). Calcium signaling. *Cell* 131, 1047–1058. doi: 10.1016/j.cell.2007.11.028

Cole, N., Brown, M., and Phillips, W. (2001). Genetic × environment interactions on blood constituents of Angus, Brahman, and reciprocal-cross cows and calves grazing common bermudagrass or endophyte-infected tall fescue. *J. Anim. Sci.* 79, 1151–1161. doi: 10.2527/2001.7951151x

Cortes, A., Maksymowych, W. P., Wordsworth, B. P., Inman, R. D., Danoy, P., Rahman, P., et al. (2015). Association study of genes related to bone formation and resorption and the extent of radiographic change in ankylosing spondylitis. *Ann. Rheum. Dis.* 74, 1387–1393. doi: 10.1136/annrheumdis-2013-204835

Dehghan, A., Kottgen, A., Yang, Q., Hwang, S. J., Kao, W. L., Rivadeneira, F., et al. (2008). Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* 372, 1953–1961. doi: 10.1016/s0140-6736(08)61343-4

Feher, J. (2017). "Calcium and phosphorus homeostasis I: the calcitropic hormones," in *Quantitative Human Physiology*, 2nd Edn, ed. J. Feher (Boston: Academic Press), 924–932. doi: 10.1016/b978-0-12-800883-6.00090-2

Feng, X., Scott, A., Wang, Y., Wang, L., Zhao, Y., Doerner, S., et al. (2014). PTPRT regulates high-fat diet-induced obesity and insulin resistance. *PLoS One* 9:e100783. doi: 10.1371/journal.pone.0100783

Fernando, R., Toosi, A., Wolc, A., Garrick, D., and Dekkers, J. (2017). Application of whole-genome prediction methods for genome-wide association studies: a Bayesian approach. *J. Agric. Biol. Environ. Stat.* 22, 172–193. doi: 10.1007/s13253-017-0277-6

Frankel, T. L., Seshadri, M. S., Mcdowall, D. B., and Cornish, C. J. (1986). Hypervitaminosis A and calcium-regulating hormones in the rat. *J. Nutr.* 116, 578–587. doi: 10.1093/jn/116.4.578

Gannagé-Yared, M.-H., Makrythanasis, P., Chouery, E., Sobacchi, C., Mehawej, C., Santoni, F. A., et al. (2014). Exome sequencing reveals a mutation in DMP1 in a family with familial sclerosing bone dysplasia. *Bone* 68, 142–145. doi: 10.1016/j.bone.2014.08.014

Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32, 361–369. doi: 10.1002/gepi.20310

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136

Gesellschaft für Ernährungsphysiologie (2006). *Empfehlungen zur Energie-und Nährstoffversorgung beim Schwein. Ausschuss für Bedarfsnormen der Gesellschaft für Ernährungsphysiologie*. Frankfurt: DLG Verlag.

Guerder, S., Hassel, C., and Carrier, A. (2018). Thymus-specific serine protease, a protease that shapes the CD4 T cell repertoire. *Immunogenetics* 71, 223–232. doi: 10.1007/s00251-018-1078-y

Hendriks, W. J. A. J., Elson, A., Harroch, S., Pulido, R., Stoker, A., and Den Hertog, J. (2013). Protein tyrosine phosphatases in health and disease. *FEBS J.* 280, 708–730. doi: 10.1111/febs.12000

Hiersche, M., Rühle, F., and Stoll, M. (2013). Postgwas: advanced GWAS interpretation in R. *PLoS One* 8:e71775. doi: 10.1371/journal.pone.0071775

Hittmeier, L. J., Grapes, L., Lensing, R. L., Rothschild, M. F., and Stahl, C. H. (2006). Genetic background influences metabolic response to dietary phosphorus restriction. *J. Nutr. Biochem.* 17, 385–395. doi: 10.1016/j.jnutbio.2005.08.008

Hoffmann, W. E., and Solter, P. (2008). "Diagnostic enzymology of domestic animals," in *Clinical Biochemistry of Domestic Animals*, 6th Edn, eds J. Kaneko, J. Harvey, and M. Bruss (Burlington, MA: Elsevier), 351–378. doi: 10.1016/b978-0-12-370491-7.00012-x

Hunter, D., Lange, M., Snieder, H., Macgregor, A., Swaminathan, R., Thakker, R., et al. (2002). Genetic contribution to renal function and electrolyte balance: a twin study. *Clin. Sci.* 103, 259–265. doi: 10.1042/cs1030259

Jeffreys, H. (1998). *The Theory of Probability*. New York, NY: Oxford University Press.

Jiang, H., and Stephens, N. L. (1994). Calcium and smooth muscle contraction. *Mol. Cell. Biochem.* 135, 1–9. doi: 10.1007/BF00925956

Just, F., Oster, M., Büsing, K., Borgelt, L., Murani, E., Ponsuksili, S., et al. (2018a). Lowered dietary phosphorus affects intestinal and renal gene expression to maintain mineral homeostasis with immunomodulatory implications in weaned piglets. *BMC Genomics* 19:207. doi: 10.1186/s12864-018-4584-2

Just, F., Reyer, H., Muráni, E., Ponsuksili, S., Oster, M., and Wimmers, K. (2018b). Genetic variants of major genes contributing to phosphate and calcium homeostasis and their association with serum parameters in pigs. *J. Appl. Genet.* 59, 1–9. doi: 10.1007/s13353-018-0449-2

Karkkainen, H. P., and Sillanpaa, M. J. (2012). Robustness of Bayesian multilocus association models to cryptic relatedness. *Ann. Hum. Genet.* 76, 510–523. doi: 10.1111/j.1469-1809.2012.00729.x

Kestenbaum, B., Glazer, N. L., Köttgen, A., Felix, J. F., Hwang, S.-J., Liu, Y., et al. (2010). Common genetic variants associate with serum phosphorus concentration. *J. Am. Soc. Nephrol.* 21, 1223–1232. doi: 10.1681/ASN.2009111104

Koyama, T., Matsui, D., Kuriyama, N., Ozaki, E., Tanaka, K., Oze, I., et al. (2015). Genetic variants of SLC17A1 are associated with cholesterol homeostasis and hyperhomocysteinaemia in Japanese men. *Sci. Rep.* 5:15888. doi: 10.1038/srep15888

Krols, M., Bultynck, G., and Janssens, S. (2016). ER–Mitochondria contact sites: a new regulator of cellular calcium flux comes into play. *J. Cell Biol.* 214, 367–370. doi: 10.1083/jcb.201607124

Kyriakides, T. R., Zhu, Y.-H., Smith, L. T., Bain, S. D., Yang, Z., Lin, M. T., et al. (1998). Mice that lack thrombospondin 2 display connective tissue abnormalities that are associated with disordered collagen fibrillogenesis, an increased vascular density, and a bleeding diathesis. *J. Cell Biol.* 140, 419–430. doi: 10.1083/jcb.140.2.419

Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11, 241–247. doi: 10.1038/ng1195-241

Lederer, E. (2014). Regulation of serum phosphate. *J. Physiol.* 592, 3985–3995. doi: 10.1113/jphysiol.2014.273979

Li, H. (2011). A quick method to calculate QTL confidence interval. *J. Genet.* 90, 355–360. doi: 10.1007/s12041-011-0077-7

Limviphuvadh, V., Chua, L. L., Eisenhaber, F., Adhikari, S., and Maurer-Stroh, S. (2010). Is LGI2 the candidate gene for partial epilepsy with pericentral spikes? *J. Bioinform. Comput. Biol.* 8, 117–127. doi: 10.1142/S0219720010004550

Liu, M., Xu, P., O'brien, T., and Shen, S. (2017). Multiple roles of Ulk4 in neurogenesis and brain function. *Neurogenesis* 4:e1313646. doi: 10.1080/23262133.2017.1313646

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Morito, N., Yoh, K., Usui, T., Oishi, H., Ojima, M., Fujita, A., et al. (2018). Transcription factor MafB may play an important role in secondary hyperparathyroidism. *Kidney Int.* 93, 54–68. doi: 10.1016/j.kint.2017.06.023

Nielson, C. M., Zmuda, J. M., Carlos, A. S., Wagoner, W. J., Larson, E. A., Orwoll, E. S., et al. (2012). Rare coding variants in ALPL are associated with low serum alkaline phosphatase and low bone mineral density. *J. Bone Miner. Res.* 27, 93–103. doi: 10.1002/jbmr.527

Pérez, P., and de Los Campos, G. (2014). Genome-wide regression & prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442

Pike, J. W., Meyer, M. B., Martowicz, M. L., Bishop, K. A., Lee, S. M., Nerenz, R. D., et al. (2010). Emerging regulatory paradigms for control of gene expression by 1,25-dihydroxyvitamin D3. *J. Steroid Biochem. Mol. Biol.* 121, 130–135. doi: 10.1016/j.jsbmb.2010.02.036

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337. doi: 10.1093/bioinformatics/btq419

Reiner, G., Clemens, N., Fischer, R., Köhler, F., Berge, T., Hepp, S., et al. (2009). Mapping of quantitative trait loci for clinical–chemical traits in swine. *Anim. Genet.* 40, 57–64. doi: 10.1111/j.1365-2052.2008.01804.x

Reyer, H., Hawken, R., Murani, E., Ponsuksili, S., and Wimmers, K. (2015). The genetics of feed conversion efficiency traits in a commercial broiler line. *Sci. Rep.* 5: 16387. doi: 10.1038/srep16387

Rodehutscord, M. (2001). Current phosphorus evaluation systems for livestock in Germany. *Lohmann Inform.* 25, 1–8.

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Shaker, J. L., and Deftos, L. (2018). "Calcium and phosphate homeostasis," in *Endotext*, eds L. J. De Groot, G. Chrousos, K. Dungan, K. R. Feingold, A. Grossman, J. M. Hershman, et al. (South Dartmouth, MA: MDText. com, Inc.).

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* 54, 1.30.1–1.30.33. doi: 10.1002/cpbi.5

Sutherland, D. J., Pujic, Z., and Goodhill, G. J. (2014). Calcium signaling in axon guidance. *Trends Neurosci.* 37, 424–432. doi: 10.1016/j.tins.2014.05.008

Tizioto, P. C., Decker, J. E., Taylor, J. F., Schnabel, R. D., Mudadu, M. A., Silva, F. L., et al. (2013). Genome scan for meat quality traits in Nelore beef cattle. *Physiol. Genomics* 45, 1012–1020. doi: 10.1152/physiolgenomics.00066.2013

Van Goor, A., Ashwell, C. M., Persia, M. E., Rothschild, M. F., Schmidt, C. J., and Lamont, S. J. (2016). Quantitative trait loci identified for blood chemistry components of an advanced intercross line of chickens under heat stress. *BMC Genomics* 17:287. doi: 10.1186/s12864-016-2601-x

van Haelst, M. M., Maiburg, M., Baujat, G., Jadeja, S., Monti, E., Bland, E., et al. (2008). Molecular study of 33 families with Fraser syndrome new data and mutation review. *Am. J. Med. Genet. A* 146A, 2252–2257. doi: 10.1002/ajmg.a.32440

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Yokoyama, K. (2008). The calcium–phosphorus in guidelines for CKD-MBD. *Kidney Int.* 74:245. doi: 10.1038/ki.2008.175

Yoo, C.-K., Cho, I.-C., Lee, J.-B., Jung, E.-J., Lim, H.-T., Han, S.-H., et al. (2012). QTL analysis of clinical-chemical traits in an F2 intercross between Landrace and Korean native pigs. *Physiol. Genomics* 44, 657–668. doi: 10.1152/physiolgenomics.00172.2011

Yuasa, T., Kataoka, H., Kinto, N., Iwamoto, M., Enomoto-Iwamoto, M., Iemura, S.-I., et al. (2002). Sonic hedgehog is involved in osteoblast differentiation by cooperating with BMP-2. *J. Cell. Physiol.* 193, 225–232. doi: 10.1002/jcp.10166

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Reyer, Oster, Wittenburg, Murani, Ponsuksili and Wimmers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 02 July 2019
doi: 10.3389/fvets.2019.00214

Transcriptomic Profiling of Duodenal Epithelium Reveals Temporally Dynamic Impacts of Direct Duodenal Starch-Infusion During Dry Period of Dairy Cattle

*Cong-Jun Li[1], Shudai Lin[1,2], María Jose Ranilla-García[3] and Ransom L. Baldwin VI[1]**

[1] Animal Genomics and Improvement Laboratory, Agricultural Research Service, Beltsville Agricultural Research Center, USDA, Beltsville, MD, United States, [2] Department of Animal Genetics, Breeding and Reproduction, College of Animal Science, South China Agricultural University, Guangzhou, China, [3] Departamento de Producción Animal, Instituto de Ganadería de Montaña, CSIC-Universidad de León, Campus de Vegazana, León, Spain

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
Mick Watson,
University of Edinburgh,
United Kingdom
Michael Mullen,
Athlone Institute of
Technology, Ireland

***Correspondence:**
Ransom L. Baldwin VI
ransom.baldwin@ars.usda.gov

Specialty section:
This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Veterinary Science

Received: 21 September 2018
Accepted: 14 June 2019
Published: 02 July 2019

Citation:
Li C-J, Lin S, Ranilla-García MJ and Baldwin RL VI (2019) Transcriptomic Profiling of Duodenal Epithelium Reveals Temporally Dynamic Impacts of Direct Duodenal Starch-Infusion During Dry Period of Dairy Cattle.
Front. Vet. Sci. 6:214.
doi: 10.3389/fvets.2019.00214

Previous research has demonstrated a positive relationship between dietary Metabolisable Energy Intake (MEI) and increased maintenance energy costs associated with the visceral tissues. Limitations in understanding this relationship include a lack of access to samples to assess regulatory control of the putative response gastrointestinal tissues to nutrients. This experiment was conducted with a single nutrient (starch hydrolysate) infused (7 d) directly into the intestine to mimic typical changes in post-ruminal starch delivery in dairy production settings. Duodenal epithelial samples collected via biopsy were evaluated using next-generation sequencing technology (RNA-Seq) to validate the use of this approach for the profiling and comparison of the transcriptome of cattle intestinal epithelial tissues. Samples of intestinal epithelial tissue were collected prior to and during the infusion of starch hydrolysate. Biopsies were collected on day 0 before and day 1, day 3, and day 7 during the infusion. Additionally, samples were collected on day 1 and day 7 after infusion was discontinued (Day 8 and Day14 of the experiment). Evaluation of RNA-seq data revealed dynamic changes in global gene expression during infusion. On day 7 of the infusion, 1490 genes were found to be differentially expressed (DE) compared to the day 0 control samples with FDR $p < 0.05$, vs. 105 genes on day 1 and 246 genes on Day 3. However, on day 8, after infusion was terminated for 24 h, only 428 genes were identified as differentially expressed compared to day 0 and only 107 genes continued to be identified by Day 14. Thus, the apparent differential expression of these genes is putatively a result of the single nutrient infused. Further, performing function and pathway analysis of the identified DE genes using IPA, we observe changes in digestive system development, and function pathways are among the primary functions of the DE genes, as well as immune response elements. Finally, primary transcription regulators such as PTH, JUN, WNT, and TNFRSF11B

were identified as the activated upstream regulators for specific future focus. Using a serial biopsy approach we are able to identify differentially expressed genes from cow duodenal epithelial tissue in response to a short-term perturbation with infused starch hydrolysate.

Keywords: ruminant, dairy, duodenal, RNA seq analysis, starch, dry period, biopsy

INTRODUCTION

Due to their central role in the absorption, processing, and assimilation of nutrients combined with relatively high metabolic activities, digestive tract tissues, and liver greatly influence maintenance energy requirements (1). Feed costs consistently represent >50% of production costs for dairy (2), thus incremental gains in feed efficiency through diet and ration management have a potentially large economic impact on producers. A great deal of research interest has been focused on these organs in ruminants. In fact, they have been shown to be affected by changes in metabolizable energy (ME) intake (1), protein intake (3, 4), nutrient restriction and/or realimentation (1, 6), and energy density of the diet (5, 6). Changes in mass associated with physiological state, when dietary energy intake is maintained, have been equivocal (7). Interestingly, it has been demonstrated (6) in growing steers that growth of the small intestine, liver, and forestomachs was the result of different processes following realimentation (hyperplastic growth, hypertrophic growth, and both, respectively). Small intestinal growth responses generally appear to be due to increases in cell number across a variety of dietary treatments, including nutrient restriction (4, 6).

The response of the rumen and small intestine to increased physiological demand for nutrients appears to be due largely to increased mass because of cell proliferation (4). Use of proliferation indices, including BrdU incorporation, Ki67 antigen staining, and tritiated thymidine incorporation assays, do not appear to be sufficiently sensitive to allow for accurate prediction of tissue proliferation status in the lactating dairy cow (8). Given that transit time through the cell cycle in the intestine is short (2–3 d); (9), a small change in transit time through the cell cycle can elicit a large net effect on total tissue proliferation. The specific molecular mechanisms regulating this increase in intestinal mass are not well studied in ruminants largely due to a lack of repeated access to these tissues. Using duodenally cannulated dairy cows we are now able to obtain serial biopsies from the duodenum using gastrointestinal endoscopy tools concomitantly with direct delivery of partially hydrolyzed starch to mimic increased post-ruminal delivery of starch from a ration. Using this approach we can procure samples to elucidate the ontogeny of the transcriptomic responses to increased starch.

The transcriptome is the essential and functional part of the genome (10). In this study, using transcriptomics and bioinformatics, we have compiled an information rich dataset and identified a large number of candidate genes for future experimental focus. Moreover, for the first time, we have identified specific transcriptomic regulators and identified pathways altered by direct infusion of a single nutrient (starch) using real-time transcriptomic profiles of the duodenal epithelium.

MATERIALS AND METHODS

All animal procedures were conducted under the approval of the Beltsville Location Institutional Animal Care and Use Committee (Protocol #15-008).

Animals, Treatments, and Sampling

Six multiparous Holstein cows fitted with duodenal and ruminal cannulae were sampled during the dry period. The cows were fed standard diets *ad libitum* as a Total Mixed Ration (TMR; 50% corn silage and 50% concentrate at a dry matter basis) with free access to fresh water. Ruminal cannulas purchased (10.2 cm interior diameter; Bar Diamond, Inc., Parma, ID) and T-shaped duodenal cannulas (i.d., 2.5 cm; Tygon tubing R-3603; Norton Co., Akron, OH) were made by fusing with cyclohexanone.

TABLE 1 | Profiling of transcriptome and responses of the intestinal epithelium to starch hydrolysate direct infusion[a].

Time point	Total gen transcripts detected	Number of degs[b]	Up-regulated gene	Down-regulated gene
D1	15,643	105	44	61
D3	15,646	245	159	86
D7	16,845	1,490	851	639
D8	16,839	428	270	158
D14	16,708	107	25	82

[a] *In comparison to pre-infusion at 0 h, genes were identified to be impacted in the intestinal epithelium by partially hydrolysed starch infusion at a stringent cutoff of FDR < 0.01.*
[b] *DEGs: differentially expressed genes.*

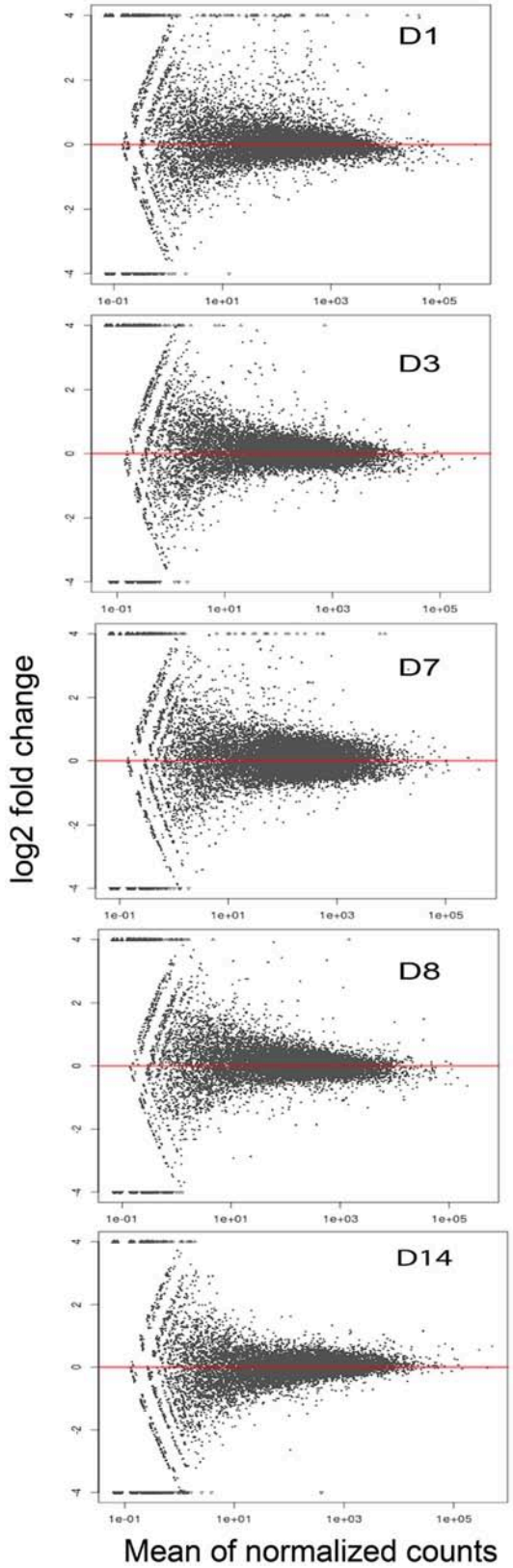


FIGURE 1 | RNA-seq MA Plots for mean six normalized samples show differential expression of genes at five sampling time points.

Duodenal cannulas were placed approximately 15-cm distal to the pylorus.

Prior to initiating infusions and sampling, cows were moved to a tie stall barn for adaptation and acclimation for at least 5 days prior to the infusion experiment. Infusion of a partially hydrolyzed starch solution (to achieve 20% MEI coming from infusate) was initiated immediately following 0 h sampling and thereafter continued for 168 h (7 days) at a rate of 5.0 L/d of a corn starch solution prepared as described by Bauer et al. (11) and stored (−20°C) until infused. After 168 h infusion, cows were maintained on the standard ration without starch infusion for an additional 168 h. Duodenal biopsies (20–30 mg/biopsy) were serially collected at 0, 24, 72, and 168 h of infusion, and 24 and 168 h post infusion through the duodenal cannula using sterile biopsy forceps aided by a Pentax EC-383IL colonoscope (PENTAX of America, New Jersey, 07645-1782 USA). Biopsies were rinsed in saline and placed into RNAlater and handled per manufacturer recommendations. Samples were stored frozen (−80°C) until sequencing.

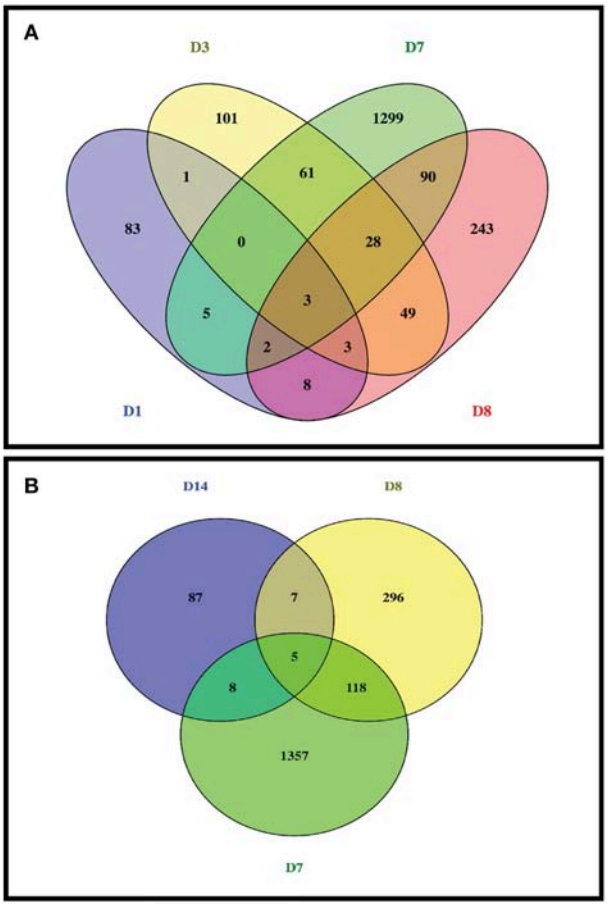


FIGURE 2 | Venn diagram of diferentially expressed genes induced by partially hydrolysed starch infusion at different sampling points. **(A)** Sample points of D1, D3, D7, and D8; **(B)** Sample time points of D7, D8, and D14.

RNA Sequencing and Bioinformatic Analysis

RNA-sequencing: Samples were isolated using Qiagen RNeasy Plus Mini Kit (Qiagen). The quality check was performed using Tapestation RNA HS Assay (Agilent Technologies, CA, USA) and quantified by Qubit RNA HS assay (ThermoFisher). Ribosomal RNA depletion was performed with Ribo-zero Magnetic Gold Kit (Catalog number MRZG12324, Illumina Inc., San Diego, CA. Samples are randomly primed and fragmented based on manufacturer's recommendation (NEBNext® Ultra™ RNA Library Prep Kit for Illumina®). The first strand is synthesized with the Protoscript II Reverse Transcriptase with a more extended extension period (40 min for 42°C). All remaining steps for library construction were used according to the NEBNext® Ultra™ RNA Library Prep Kit for Illumina®. Illumina 8-nt dual-indices were used. Samples were pooled and sequenced on a HiSeq with a read length configuration of 150 PE.

Bioinformatic Analysis: Raw data quality assessment and preprocessing. During the library preparation and sequencing, artificial/technical biases, as well as sample contamination, could be introduced and affect the accuracy of the downstream statistical analysis (Mapping statistics are provided in **Supplementary Table 1**). We performed a thorough quality assessment using FASTQC (version v0.11.3). Sequence alignment: STAR (version 2.5.2b) (12), a splice aware aligner, was used to perform the RNA-Seq alignment. The UMD3.1 and UMD3.1.90 from Ensembl were used as genome reference and annotation reference, respectively during the alignment. Then dupRader and Picard CollectRnaSeqMetrics (version 2.10.5) were used to evaluate duplicates level and overall alignment performance.

Gene Expression Estimation and Differential Expression Analysis: We used HT-Seq (version 0.6.0) (13) to calculate the per gene expression count and DE-Seq (14) was used to estimate the differentially expressed genes. Some quality control assessments, as well as downstream exploratory analysis, were primarily performed using R package including but not limited to mixOmics, clusterProfiler, topGO, DOSE, pathview, and org.Bt.eg.db.

Functional Annotation of Differentially Expressed Genes: Ingenuity Pathways Analysis (IPA, Qiagen) was used to further identify the molecular processes, molecular functions, and genetic networks affected by starch infusion through analysis of the identified differentially expressed genes. As an integrated analysis software, IPA is a software application that enables users to identify the biological mechanisms, pathways, and functions most relevant to their experimental datasets or genes of interest. The "core analysis" function included in the IPA software was used to interpret the

TABLE 2 | Top GO terms in biological processes significantly impacted temporally by partially hydrolysed starch infusion[a,b].

GO ID	Description ion	Gene ratio	BgRatio	*P*-value	*P*-value adjusted	*q*-value
D3						
GO:0019221	Cytokine-mediated signaling pathway	8/58	76/5,527	9.7×10^{-7}	0.0009	0.0008
GO:0071345	Cellular response to cytokine stimulus	9/58	106/5,527	1.21×10^{-6}	0.0009	0.0008
GO:0034097	Response to cytokine	9/58	122/5,527	3.96×10^{-6}	0.0020	0.0017
GO:0046627	Negative regulation of insulin receptor signaling pathway	3/58	7/5,527	3.73×10^{-5}	0.0131	0.0107
GO:1901700	Response to oxygen-containing compound	11/58	257/5,527	5.76×10^{-5}	0.0131	0.0107
D7						
GO:0008150	Biological process	421/421	4,837/5,527	3.72×10^{-26}	< 0.0001	< 0.0001
GO:0008152	Metabolic process	317/421	3,464/5,527	7.17×10^{-9}	< 0.0001	< 0.0001
GO:0009987	Cellular process	360/421	4,158/5,527	6.23×10^{-8}	< 0.0001	< 0.0001
GO:0044238	Primary metabolic process	256/421	2,843/5,527	3.64×10^{-5}	0.0263	0.0258
GO:0071704	Organic substance metabolic process	263/421	2,970/5,527	1.03×10^{-4}	0.0428	0.0420
D8						
GO:0008150	Biological process	125/125	4,837/5,527	4.7×10^{-8}	< 0.0001	< 0.0001
D14						
GO:0065007	Biological regulation	24/27	2,624/5,527	7.79×10^{-6}	0.0059	0.0049
GO:0050789	Regulation of biological process	23/27	2,445/5,527	1.33×10^{-5}	0.0059	0.0049
GO:0050794	Regulation of cellular process	21/27	2,257/5,527	1.01×10^{-4}	0.0302	0.0248
GO:0050896	Response to stimulus	18/27	1,832/5,527	3.64×10^{-4}	0.0499	0.0411
GO:0007094	Mitotic spindle assembly checkpoint	2/27	8/5,527	6.32×10^{-4}	0.0499	0.0411

[a] *GO: gene ontology.*

[b] *All the time points (day) are compared against D 0 (baseline control); 2. Gene Ratio = the number of all genes assigned to this GO term to the number of significantly regulated genes that can be assigned to this GO term, BgRatio: ratio between the number of genes in the pathway and the total examined background of genes. P-value: for hypergeometric test; P-value adjusted: P-value for hypergeometric test adjusted for Benjamini-Hochberg correction.*

differentially expressed data, which included identification of probable biological processes, canonical pathways, upstream transcriptional regulators, and gene networks responding to the starch infusion. The temporally dynamic changes in gene activities during starch infusion were also compared using IPA.

RESULTS

RNA-seq Revealed Dynamic Changes in Global Gene Expression of Cattle Intestinal Epithelium During Infusion

From RNA sequencing reads of 30 intestinal epithelial samples (6 animals with 5 sampling time points on Day 0 (D0), Day 1 (D1), Day 3 (D3), Day 7 (D7), and Day 14 (D14)), a total of varied from 15,643 to 16,845 genes were detected from at least one of the RNA sequenced samples, and highest number of genes (16,845) was detected in D7 of starch hydrolysate sampling. Likewise, total gene transcripts detected attained apeak on D7 (**Table 1**). In comparison to pre-infusion at 0 h, a total of 1,795 DE genes were identified at least once at the different sampling time points at a stringent cutoff of FDR <0.01 as impacted in the biopsies of intestinal epithelium in response to the starch infusion. The apparent maximal effect of starch hydrolysate infusion was observed on day 7 (**Table 1**; **Figure 1**) where 1,490 genes were found to be differentially expressed (DEGs) with FDR $p < 0.05$, compared to 105 genes on D1 and 246 genes on D3. After the partially hydrolyzed starch infusion was terminated, on D8 (one-day post infusion), only 428 genes were identified differentially expressed compared to day 0 and only107 genes on D14 (7-day post-infusion). While numerous genes are overlapping across different time points among these impacted genes, 78 genes were responsive only on D1, 148 genes were only on D3, and 1,380 genes were impacted only on D7. This occurred in such a way that most of the impacted genes are only represented at one or two sampling points. The overlapping and specific responding genes at the different sampling points were illustrated in a Venn diagram (**Figure 2A**).

Gene Ontology (GO) Enrichment Analysis of Differentially Expressed Genes Impacted by Starch Infusion

Gene ontology (GO) enrichment analysis of the differentially expressed genes was performed to further clarify the putative functions affected by starch infusion within each sampling day compared to control samples from Day 0. Changes of GO terms from enrichment analysis across sampling days support the concept of a coordinated and dynamic response temporal response to starch infusion by the duodenal epithelial transcriptome. The top GO terms in the biological processes significantly enriched in differentially expressed genes from each sampling point are listed in **Table 2**. While there are 105 DEGs for the D1 sampling, no enriched GO terms were identifiable. By day 7 of starch infusion the most significantly enriched GO

TABLE 3 | Top enriched GO terms in molecular functions significantly impacted temporally by partially hydrolysed starch infusion[a,b].

GO ID	Description	GeneRatio	BgRatio	*P*-value	*P*-value adjusted	*q*-value
D3						
GO:0003674	Molecular function	54/54	4,363/5,527	2.65×10^{-6}	0.0006	0.0005
GO:0005126	Cytokine receptor binding	6/54	83/5,527	1.38×10^{-4}	0.0105	0.0094
GO:0042802	Identical protein binding	11/54	307/5,527	1.45×10^{-4}	0.0105	0.0094
GO:0008009	Chemokine activity	3/54	17/5,527	5.44×10^{-4}	0.0256	0.0229
GO:0005125	cytokine activity	5/54	75/5,527	7.58×10^{-4}	0.0256	0.0230
D7						
GO:0003674	Molecular function	382/382	4,363/5,527	1.44×10^{-41}	< 0.0001	< 0.0001
GO:0005488	Binding	304/382	3,268/5,527	1.33×10^{-18}	< 0.0001	< 0.0001
GO:0003824	Catalytic activity	198/382	1,872/5,527	5.37×10^{-14}	< 0.0001	< 0.0001
GO:0043167	Ion binding	179/382	1,678/5,527	1.58×10^{-12}	< 0.0001	< 0.0001
GO:0046872	Metalion binding	119/382	1,066/5,527	5.99×10^{-9}	< 0.0001	< 0.0001
D8						
GO:0003674	Molecular function	118/118	4,363/5,527	5.42×10^{-13}	< 0.0001	< 0.0001
D14						
GO:0004872	Receptor activity	7/24	224/5,527	3.14×10^{-5}	0.0050	0.0041
GO:0060089	Molecular transducer activity	7/24	293/5,527	1.74×10^{-4}	0.0107	0.0089
GO:0008329	Signaling pattern recognition receptor activity	2/24	132/4,887	2.68×10^{-4}	0.0107	0.0089
GO:0038187	Pattern recognition receptor activity	2/24	132/4,887	2.68×10^{-4}	0.0107	0.0089
GO:0038023	Signaling receptor activity	5/24	188/5,527	1.08×10^{-4}	0.0344	0.02879

[a] *GO: gene ontology.*

[b] *All the time points (day) are compared against D0 (baseline control); 2. GeneRatio = the number of all genes assigned to this GO term to the number of significantly regulated genes that can be assigned to this GO term, BgRatio: ratio between the number of genes in the pathway and the total examined background of genes. P-value: for hypergeometric test; P-value adjusted: P-value for hypergeometric test adjusted for Benjamini-Hochberg correction.*

terms included: biological processes, metabolic processes, cellular processes, primary metabolic processes, and organic substance metabolic processes. The top GO terms enriched for DEGs at each sampling point (D3, D7, D8, and D14) are presented in **Table 2**. All of the genes present for each GO terms in the dataset are listed in **Supplementary Table 2**. In addition to the biological processes, GO terms in molecular activities categories were also analyzed. Interestingly, most GO terms in molecular activity in the 36 samples were related to the following molecular functions; Cytokine receptor binding, catalytic activity; and molecular binding (**Table 3**; **Supplementary Table 3**), as we found for biological functions, on D1 there is no enriched GO term detected for molecular functions.

Functional Annotation of Differentially Expressed Genes Using IPA

To investigate further the biological functions affected by the starch infusion, Ingenuity Pathways Analysis (IPA) was utilized. Comparison analysis using IPA was performed to elucidate the dynamics and the tendency of the biological and molecular functions impacted by starch infusion through the experimental infusion period. The top affected functions of the identified DEGs at each sampling time point are presented in two figures, **Figures 3**, **4**. In **Figure 3A**, the top biological functions impacted by starch infusion are presented in a heatmap according to their activation z-scores. The predominant positive biological functions impacted during the whole experimental course were the growth of connective tissue, the growth of epithelial tissue, and proliferation of epithelial cells. Consistently, the primary physiological functions of the DEGs on D7 are digestive system development and function related (**Figure 3B**). In **Figure 4**, top molecular functions significantly impacted by starch infusion are listed. The results were consistent with the GO enrichment analysis.

In addition to the biological and molecular functions, IPA analysis revealed canonical pathways putatively affected as determined from the DEGs. Some essential canonical pathways were induced by starch-infusion such as ERK/MARK presented graphically in **Figure 5**. The heat map is used to visualize the pathway scores (Activation z-score) and the expression of the genes involved in the canonical pathways network. These heat maps can also show changes in relative expression across the five sampling time points (D1, D2, D7, D8, and D14) simultaneously. **Figure 5** presents the activation z-score of ERK/MARK singling pathway with the expression of genes in ERK/MARK signaling pathway network. The activation z-score for this pathway is at its highest activation status on the D3, D7 which continues to D8 of starch infusion.

Using IPA analysis, potential upstream regulators of the DEGs in response to starch infusion were identified. Upstream regulators of the DEGs in this data set are identified based on known molecular actions from the literature and thus, may be involved with regulation of the response observed in the DEG identified. The top five upstream regulators at the different sampling time points are listed in **Table 4**. There is apparent overlap in the regulatory actions of these regulators

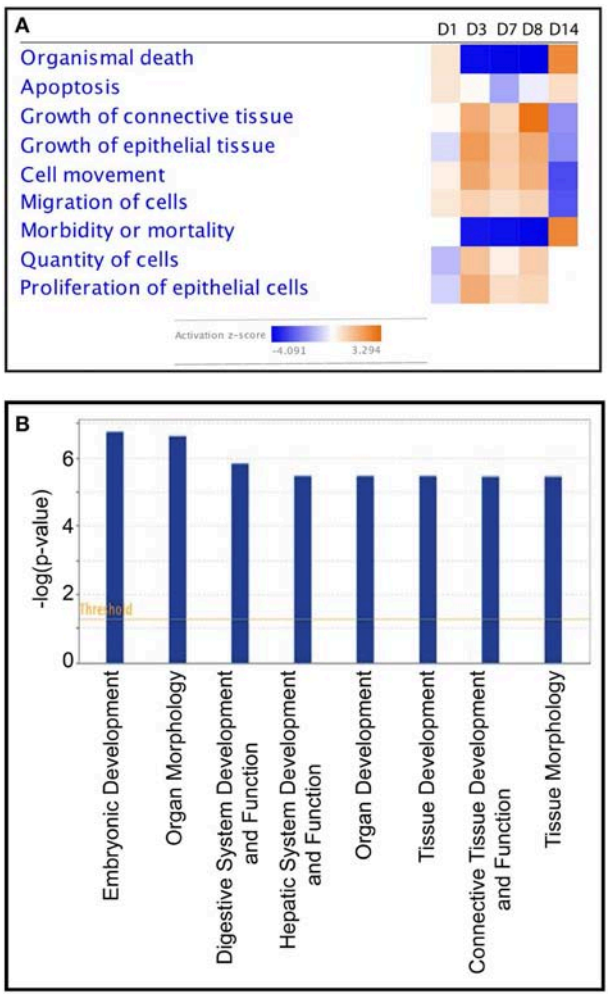


FIGURE 3 | (A) Heatmap—comparison of the top biological functions impacted by partially hydrolysed starch infusion at the different samspling time points with activation z-score. **(B)** Top physiological functions of DEGs on D7.

with the affected cellular functions as illustrated in **Figure 6**. On D3, upregulated upstream regulators FOXM1 and AREG are overlapped with upregulated genes of CCNF, MKI67, AURKB, CCND1, IKBKB, STAT3, and ELF3 and the result would be effects on cell cycle progression and cell survival activities. Similarly, on D7, enhanced activities of upstream regulators PTH and CHUK would be expected to result in activation of cell cycle progression and development of epithelial tissue.

As mentioned above, during the starch infusion, the most DEGs altered in duodenal epithelium were observed in the D7 samples. Functional network analysis identifies the biologically relevant networks based on the DEGs in response to starch infusion. The top biologically relevant network on D3 is associated with lipid metabolism, molecular transport, small molecule biochemistry (**Figure 7A**) and the top biologically relevant network on D7 is associated with the biological functions of carbohydrate metabolism, lipid metabolism, and molecular transport (**Figure 7B**).

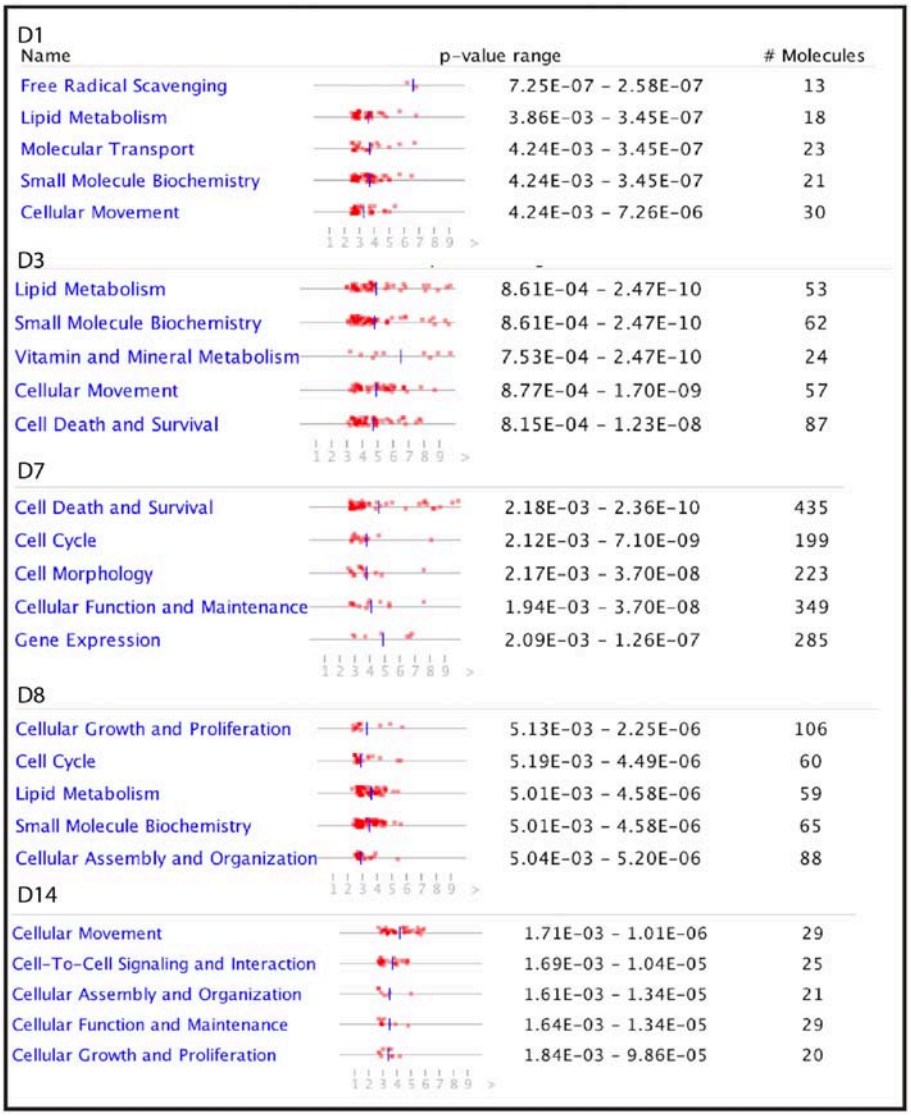


FIGURE 4 | The top five molecular functions impacted by partially hydrolysed starch infusion at the different sampling points during the experiment.

DISCUSSION

The duodenal epithelium is a highly metabolically active tissue due to the functions it performs (absorption, transport and protection). In fact, total gastrointestinal tissues use a disproportionate amount of the energy used by the animal (about 25% of total oxygen consumption) given its relative size (about 6% of body weight). Additionally understanding the extent to which individual nutrients are used by gut tissues is important to assess net nutrient needs of the animal (15). Ruminal and abomasal starch hydrolysate infusions have been used previously to study the metabolism (11) and gene expression of intestinal epithelia (16–18). However, those studies were limited by the availability of technology at the time, and only a few genes were examined. In this report, coupling next generation sequencing and transcriptomic profiling approaches with a serial biopsy sampling scheme were used to investigate the changes in the golobal transcriptome during an adaptation of intestinal epithelia of dairy cattle to a single nutrient, starch hydrolysate, by direct duodenal infusion.

The transcriptome is known to have distinct profiles unique to cell type, developmental stages, and health status (10). RNA-sequencing (RNA-seq) has been widely used as a highly reliable tool for unbiased analysis of transcriptome changes within cells and tissues (19). Using a direct biopsy technique aided by a Pentax EC-383IL colonoscope, we were also able to serially collect the duodenal epithelial samples throughout a single nutrient infusion experimental protocol lasting 14 day. Next we assembled the transcriptome and compared gene expression patterns and thus, can assess if temporal impacts on the duodenal epithelial

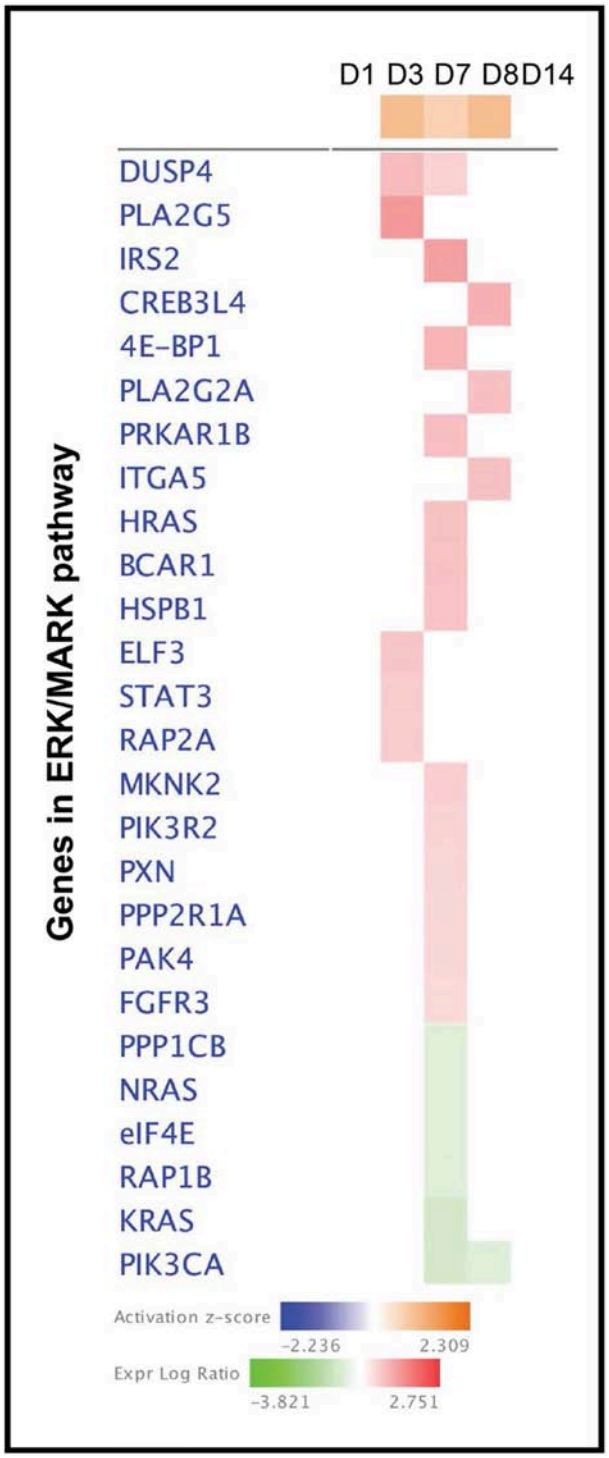


FIGURE 5 | A heatmap shows the expression of the genes in ERK/MARK pathway induced by partially hydrolysed starch infusion.

TABLE 4 | Top upstream regulators and *P*-value of overlap predicted activation.

Upstream regulator	Molecule type	*P*-value of overlap
D1		
TNF	Cytokine	3.90×10^{-09}
Cg	Complex	2.71×10^{-07}
IFNG	Cytokine	4.50×10^{-06}
EGF	Growth factor	9.76×10^{-06}
CDX1	transcription regulator	3.58×10^{-05}
D3		
CFTR	Ion channel	2.46×10^{-09}
PPARA	Ligand-dependent nuclear receptor	4.62×10^{-09}
TLR4	Transmembrane receptor	1.39×10^{-08}
MET	Kinase	2.99×10^{-08}
TNF	Cytokine	9.90×10^{-08}
D7		
MIR17HG	Other	1.39×10^{-06}
INS	Other	3.26×10^{-05}
THRB	Ligand-dependent nuclear receptor	8.61×10^{-05}
EPAS1	Transcription regulator	1.76×10^{-04}
IGFBP7	Transporter	2.17×10^{-04}
D8		
ERBB2	Kinase	1.22×10^{-11}
AREG	Growth factor	2.37×10^{-07}
PPARA	Ligand-dependent nuclear receptor	4.56×10^{-07}
TGFB1	Growth factor	6.73×10^{-07}
EP400	Other	3.26×10^{-06}
D14		
TNF	Cytokine	1.54×10^{-08}
TGFB1	Growth factor	1.85×10^{-06}
STAT3	Transcription regulator	5.43×10^{-06}
GLI2	Transcription regulator	6.17×10^{-06}
TP73	Transcription regulator	8.09×10^{-06}

transcriptome induced by starch hydrolysate infusion in dairy cattle are detectable.

Notably, the transcriptomic response occurred in a pattern where a majority of the DEG only represented at one or two sampling points used, potentially indicating a coordinated temporal pattern of changes by intestinal epithelial transcriptome induced by starch hydrolysate. After 7 d of the infusion, 1,490 genes were identified as differentially expressed (with FDR $p < 0.05$), compared to only 105 genes on D1 and 246 genes on D3. Moreover, after terminating the infusion for a day, (D8; one day post infusion), a maked decrease in DEG (only 428 DEGs) was observed compared to D0 and by D14 (7 d post termination fo infusion) only 107 genes were different from D0. Thus, it appears that differential expression of genes during infusion is putatively the result of the starch hydrolysate infusion. Mechanistically we are not able to completely rule out that other physical or environmental factors changed over time, however, the experimental design minimized other factors by maintaining cows on a consistent TMR throughout the experiment.

Regulation of gene expression within the intestinal epithelium, as with other tissues, is complex and controlled by various signaling pathways that regulate the balance between proliferation and differentiation (9). We have previously

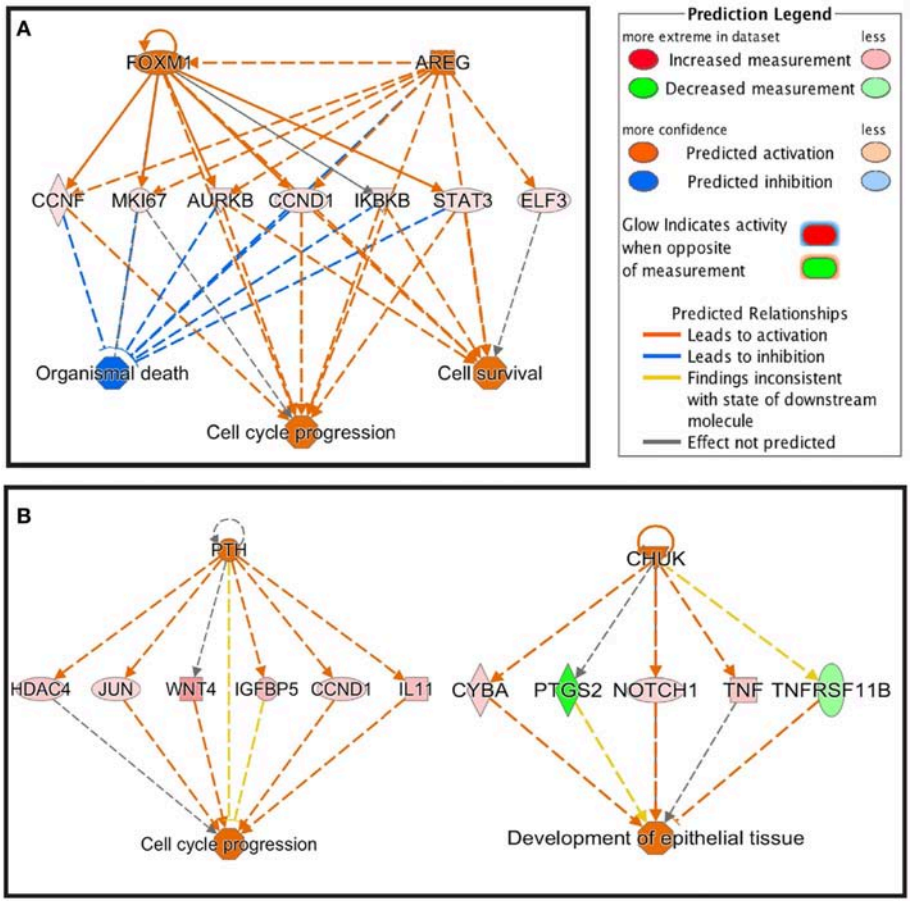


FIGURE 6 | Regulator effects of FOXM1 and AREG on D3 **(A)** and PTH and CHUK on D7 **(B)**.

identified in sheep nutrient use efficiency and body composition experiments that when nutrient density is increased (increased concentrate) by altering forage and concentration ratio in the ration, there is an increase in intestinal epithelial cell mass (5). Indeed, the DEG observed in the current experiment are predictive of an increased cell proliferation in response to the starch. Positively affected biological functions identified as impacted during the infusions were growth of connective tissue, the growth of epithelial tissue, and proliferation of epithelial cells (**Figure 3**). Clearly, these biological functions are consistent with the major molecular functions induced by starch infusion, and given the nature of the treatment, appear to be treatment specific.

Consistently, the genes in the ERK (extracellular-regulated kinase)/MAPK (mitogen activated protein kinase) signaling pathways are activated by the infusion protocol used. The ERK/MAPK pathway is a crucial pathway that transduces cellular information on meiosis/mitosis, growth, and differentiation within a cell. The mitogen-activated protein kinases (MAPK) signaling pathway is shared by four distinct cascades, including the extracellular signal-related kinases (ERK1/2), Jun amino-terminal kinases (JNK1/2/3), p38-MAPK, and ERK5 (20). ERK is also translocated into the nucleus where it induces gene transcription by interacting with transcriptional regulators like ELK-1, STAT-1 and−3, ETS, and MYC. ERK activation of p90RSK in the cytoplasm leads to its nuclear translocation where it indirectly induces gene transcription through interaction with transcriptional regulators, CREB, c-Fos, and SRF (21). This all consistent with the potential for ERK/MAPK pathway to have an important role in the epithelial response to increased luminal starch.

Our data also identified a number of immune system markers such as TNF and cytokines, which are differentially expressed after infusion of a partially hydrolyzed starch solution. This could be an indication that gut immune cells are impacted by the influx of the starch directly or may play an important role in absorption, metabolism, and transport of glucose by the epithelial tissue. The GO term analysis likewise indicates the appearance of an immune response as cytokine-mediated signaling pathway is significantly perturbed by the treatment. The gastrointestinal epithelium has a large number of immune cells integrated within the tissue presumably to enhance defense against disease-causing microbes. Recent reports have demonstrated the presence of specific types of immune cells distributed throughout the

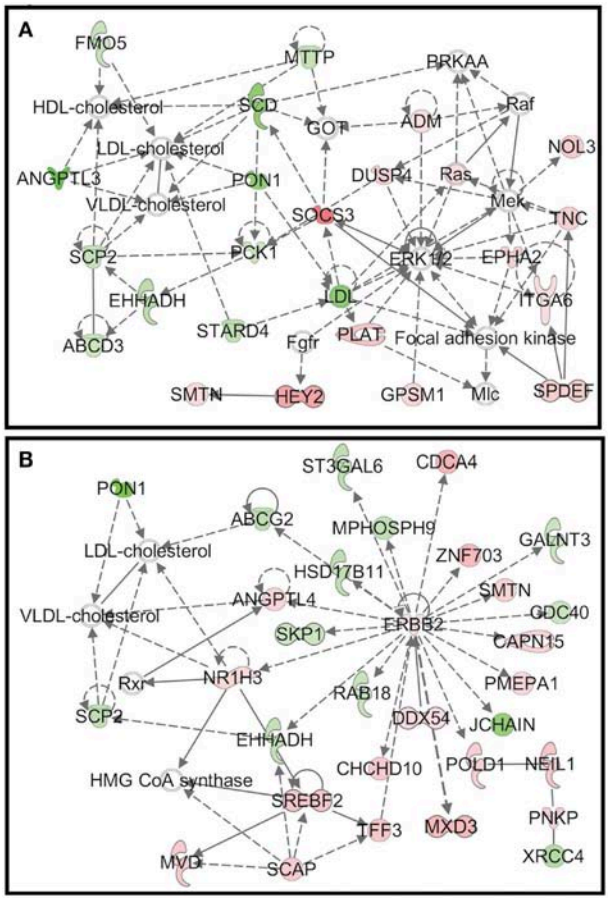


FIGURE 7 | (A) The top biologically relevant network on D3 is associated with lipid metabolism, molecular transport, small molecule biochemistry and **(B)** the top biologically relevant network on D7 is associated with the biological functions of carbohydrate metabolism, lipid metabolism, and molecular transport.

intestinal epithelium as intraepithelial lymphocytes (22, 23). They further demonstrated that in addition to their immune functions, the cells have an integral role in the control of metabolism through regulation of hormones released in response to feed consumption. This interesting finding supports the contention that immune cells may be involved in the control of metabolism. Further evidence of a metabolic role of these cells is not only their relative abundance in the sections of the intestine where nutrient absorption occurs, but also that they express genes associated with metabolism even in the absence of infection (22).

By performing function and pathway analysis of DE genes using IPA, we found, perhaps unsurprisingly, that digestive system development and function are among the primary functions of the DEGs identified. Furthermore, primary transcription regulators such as PTH, JUN, WNT, and TNFRSF11B were identified as the activated upstream regulators (**Figure 6**). Previous research provided indications that WNT signaling is important for proliferation of the intestinal epithelium (24). The enhanced activities of upstream regulators PTH and CHUK would be expected to result in activation of cell cycle progression and development of epithelial tissue. These results at the transcription level of integration demonstrate the responsive nature of the intestinal epithelial tissue to a single nutrient delivered on the luminal side with no other changes in diet. As outlined earlier, changes in gastrointestinal, and specifically epithelial, mass are known to be a response to alteration in diet and ration delivery in productive ruminants. The transciptiomic changes of these biopsied tissue are likely the necessary response to maintain epithelial homeostasis in the face of a changing nutrient supply.

Unsurprisingly given starch hydrolysate was the nutrient infused, the top functional network identified is specifically related to functions such as carbohydrate metabolism, lipid metabolism, and molecular transport (**Figure 7**). Changes in these functions as the top network reflect the temporal transcriptomic response of duodenal epithelium to the starch hydrolysate infusion. The networks also explored functional interactions among DEGs.

In summary, transcriptomic profiling with next-generation sequencing and bioinformatics were utilized to accelerate our understanding of the multiple levels of regulation ongoing in duodenal epithelial transcriptome induced by starch infusion. Use of direct infusion of a single nutrient, in combination with serial biopsy technique, has facilitated real-time sample collection and thus the ability to assess the temporal impacts on the duodenal epithelial transcriptome induced by starch hydrolysate infusion. Moreover, direct duodenal infusion of starch hydrolysate induces measurable transcriptomic responses in epithelial tissue of cattle intestine in short-term experiments that will ultimately facilitate a better understanding of the regulation of this tissue level response. Several important pathways and regulator mechaisms have been identified for future experimental focus. The use of transcriptomic profiling provides comprehensive gene expression information for improving our understanding of the molecular mechanisms involved in the intestinal functions, as well as maintaining epithelial homeostasis of cattle intestine.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Animal Care and Use Committee for the Beltsville Location. The protocol was approved by the Beltsville Location Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

C-JL oversaw sample analysis, bioinformatics, prepared figures and tables, and prepared manuscript with RB. SL participated in statistical analysis and bioinformatics. MR-G participated in experiment, sample collection, analysis of samples, and manuscript review. RB developed experimental

design, conducted all aspects of experiment, sampling, analysis, interpreting of results, and writing manuscript.

ACKNOWLEDGMENTS

Mention of a product, reagent or source does not constitute an endorsement by USDA to the exclusion of other products or services that perform a comparable function. The US Department of Agriculture is an equal opportunity provider and employer. MR-G gratefully acknowledges the receipt of a Fulbright/Ministry of Education of Spain Visiting Scholarship (FMECD-ST-2014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2019.00214/full#supplementary-material

Supplementary Table 1 | Mapping statistics of the RNA-Seq data used for profiling the transcriptome and response of the intestinal epithelium to starch hydrolysate direct infusion.

Supplementary Table 2 | GO terms (Biological Process) enriched for DEGs.

Supplementary Table 3 | GO terms (Molecular Function) enriched for DEGs.

REFERENCES

1. Johnson DE, Johnson KA, Baldwin RL. Changes in liver and gastrointestinal tract energy demands in response to physiological workload in ruminants. *J Nutr.* (1990) 120:649–55. doi: 10.1093/jn/120.6.649
2. USDA-ERS. *Milk Cost of Production Estimates*. USDA-ERS (2016). Available online at: https://www.ers.usda.gov/data-products/milk-cost-of-production-estimates.aspx (accessed June 21, 2019).
3. Villamarin B, Fernandez E, Mendez J. Analysis of grass silage from Northwestern Spain by near-infrared reflectance spectroscopy. *J AOAC Int.* (2002) 85:541–5.
4. Baldwin VIRL, McLeod KR. Effects of diet forage:concentrate ratio and metabolizable energy intake on isolated rumen epithelial cell metabolism *in vitro*. *J Anim Sci.* (2000) 78:771–83. doi: 10.2527/2000.783771x
5. McLeod KR, Baldwin RLT. Effects of diet forage:concentrate ratio and metabolizable energy intake on visceral organ growth and *in vitro* oxidative capacity of gut tissues in sheep. *J Anim Sci.* (2000) 78:760–70. doi: 10.2527/2000.783760x
6. Sainz RD, Bentley BE. Visceral organ mass and cellularity in growth-restricted and refed beef steers. *J Anim Sci.* (1997) 75:1229–36. doi: 10.2527/1997.7551229x
7. Scheaffer AN, Caton JS, Bauer ML, Reynolds L. Influence of pregnancy on body weight, ruminal characteristics, and visceral organ mass in beef heifers. *J Anim Sci.* (2001) 79:2481–90. doi: 10.2527/2001.7992481x
8. Baldwin VI RL, McLeod KR, Klotz JL, Heitmann RN. Rumen development, intestinal growth and hepatic metabolism in the pre- and postweaning ruminant. *J Dairy Sci.* (2004) 87(E. Suppl):E55–65. doi: 10.3168/jds.S0022-0302(04)70061-2
9. Richmond CA, Breault DT. Regulation of gene expression in the intestinal epithelium. *Prog Mol Biol Transl Sci.* (2010) 96:207–29. doi: 10.1016/B978-0-12-381280-3.00009-9
10. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely K, Powell J, et al. The transcriptional landscape of age in human peripheral blood. *Nat Commun.* (2015) 6:8570. doi: 10.1038/ncomms9570
11. Bauer ML, Harmon DL, McLeod KR, Huntington G. Adaptation to small intestinal starch assimilation and glucose transport in ruminants. *J Anim Sci.* (1995) 73:1828–38. doi: 10.2527/1995.7361828x
12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
13. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics.* (2015) 31:166–9. doi: 10.1093/bioinformatics/btu638
14. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* (2010) 11:R106. doi: 10.1186/gb-2010-11-10-r106
15. Britton R, Krehbiel C. Nutrient metabolism by gut tissues. *J Dairy Sci.* (1993) 76:2125–31. doi: 10.3168/jds.S0022-0302(93)77547-5
16. Liao SF, Alman MJ, Vanzant ES, Miles E, Harmon DL, McLeod KR, et al. Basal expression of nucleoside transporter mRNA differs among small intestinal epithelia of beef steers and is differentially altered by ruminal or abomasal infusion of starch hydrolysate. *J Dairy Sci.* (2008) 91:1570–84. doi: 10.3168/jds.2007-0763
17. Liao SF, Harmon DL, Vanzant ES, McLeod K, Boling JA, Matthews JC. The small intestinal epithelia of beef steers differentially express sugar transporter messenger ribonucleic acid in response to abomasal versus ruminal infusion of starch hydrolysate. *J Anim Sci.* (2010) 88:306–14. doi: 10.2527/jas.2009-1992
18. Liao SF, Vanzant ES, Harmon DL, McLeod K, Boling JA, Matthews JC. Ruminal and abomasal starch hydrolysate infusions selectively decrease the expression of cationic amino acid transporter mRNA by small intestinal epithelia of forage-fed beef steers. *J Dairy Sci.* (2009) 92:1124–35. doi: 10.3168/jds.2008-1521
19. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* (2008) 5:621–8. doi: 10.1038/nmeth.1226
20. Sun Y, Liu WZ, Liu T, Feng X, Yang N, Zhou HF, et al. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *J Recept Signal Transduct Res.* (2015) 35:600–4. doi: 10.3109/10799893.2015.1030412
21. Boglari G, Szeberenyi J. Nuclear translocation of p90Rsk and phosphorylation of CREB is induced by ionomycin in a Ras-independent manner in PC12 cells. *Acta Biol Hung.* (2002) 53:325–34. doi: 10.1556/ABiol.53.2002.3.9
22. Esterhazy D, Mucida D. Gut immune cells have a role in food metabolism. *Nature.* (2019) 566:49–50. doi: 10.1038/d41586-019-00235-5
23. He S, Kahles F, Rattik S, Nairz M, McAlpine CS, Anzai A, et al. Gut intraepithelial T cells calibrate metabolism and accelerate cardiovascular disease. *Nature.* (2019) 566:115–9. doi: 10.1038/s41586-018-0849-9
24. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters P, et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet.* (1998) 19:379–83. doi: 10.1038/1270

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Lin, Ranilla-García and Baldwin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 July 2019
doi: 10.3389/fgene.2019.00651

Co-Expression Networks Reveal Potential Regulatory Roles of miRNAs in Fatty Acid Composition of Nelore Cattle

*Priscila S.N. de Oliveira[1], Luiz L. Coutinho[2], Aline S.M. Cesar[3], Wellison J. da Silva Diniz[4], Marcela M. de Souza[5], Bruno G. Andrade[1], James E. Koltes[5], Gerson B. Mourão[3], Adhemar Zerlotini[6], James M. Reecy[5] and Luciana C.A. Regitano[1]**

[1] Embrapa Pecuária Sudeste, Empresa Brasileira de Pesquisa Agropecuária, São Carlos, Brazil, [2] Department of Animal Science, University of São Paulo, Piracicaba, Brazil, [3] Department of Agroindustry, Food and Nutrition, University of São Paulo, Piracicaba, Brazil, [4] Department of Genetics and Evolution, Federal University of São Carlos, São Carlos, Brazil, [5] Department of Animal Science, Iowa State University, Ames, IA, United States, [6] Embrapa Informática Agropecuária, Campinas, Brazil

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
Alessandra Crisà,
Council for Agricultural and Economics Research, Italy
Paul Cormican,
Teagasc Grange Animal and Bioscience Research Department (ABRD), Ireland

***Correspondence**
Luciana C.A. Regitano
luciana.regitano@embrapa.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 08 February 2019
Accepted: 19 June 2019
Published: 11 July 2019

Citation:
de Oliveira PSN, Coutinho LL, Cesar ASM, Diniz WJdS, de Souza MM, Andrade BG, Koltes JE, Mourão GB, Zerlotini A, Reecy JM and Regitano LCA (2019) Co-Expression Networks Reveal Potential Regulatory Roles of miRNAs in Fatty Acid Composition of Nelore Cattle. Front. Genet. 10:651. doi: 10.3389/fgene.2019.00651

Fatty acid (FA) content affects the sensorial and nutritional value of meat and plays a significant role in biological processes such as adipogenesis and immune response. It is well known that, in beef, the main FAs associated with these biological processes are oleic acid (C18:1 cis9, OA) and conjugated linoleic acid (CLA-c9t11), which may have beneficial effects on metabolic diseases such as type 2 diabetes and obesity. Here, we performed differential expression and co-expression analyses, weighted gene co-expression network analysis (WGCNA) and partial correlation with information theory (PCIT), to uncover the complex interactions between miRNAs and mRNAs expressed in skeletal muscle associated with FA content. miRNA and mRNA expression data were obtained from skeletal muscle of Nelore cattle that had extreme genomic breeding values for OA and CLA. Insulin and MAPK signaling pathways were identified by WGCNA as central pathways associated with both of these fatty acids. Co-expression network analysis identified bta-miR-33a/b, bta-miR-100, bta-miR-204, bta-miR-365-5p, bta-miR-660, bta-miR-411a, bta-miR-136, bta-miR-30-5p, bta-miR-146b, bta-let-7a-5p, bta-let-7f, bta-let-7, bta-miR 339, bta-miR-10b, bta-miR 486, and the genes *ACTA1* and *ALDOA* as potential regulators of fatty acid synthesis. This study provides evidence and insights into the molecular mechanisms and potential target genes involved in fatty acid content differences in Nelore beef cattle, revealing new candidate pathways of phenotype modulation that could positively benefit beef production and human consumption.

Keywords: *Bos indicus*, conjugated linoleic acid, integrative genomics, mRNA, miRNA, oleic acid

Abbreviations: CLA, conjugated linoleic acid; DH, differential hubbing; FA, fatty acid; GEBV, genomic estimated breeding value: IMF, intramuscular fat content: ME, module eigengene: MM, module membership: OA, oleic acid: PCIT, partial correlation with information theory: PIF, phenotypic impact factor: RIF1, regulatory impact factor 1: RIF2, regulatory impact factor 2: WGCNA, weighted gene co-expression network analysis:

INTRODUCTION

Fatty acid (FA) content is an important trait that can influence the sensorial and nutritional value of beef and plays a significant role in molecular and physiological processes. Despite that the consumption of beef fatty acids is being associated with metabolic diseases such as type 2 diabetes and obesity effects, such as altered blood lipid and lipoprotein content (Wood et al., 2008), beef has beneficial effects on human health due to its high nutritional value, being also an important source of oleic acid (OA) (Laaksonen et al., 2005a). Likewise, conjugated linoleic acids (CLAs) could have a range of nutritional benefits in the diet (Valsta et al., 2005).

Fatty acid biosynthesis biological processes are complex and dependent on several regulatory mechanisms, such as post-transcriptional regulation of gene expression (Nakamura and Nara, 2003). In this sense, miRNAs have been shown to block the translation of target mRNAs and thereby post-transcriptionally regulate adipogenesis and several other biological processes involved in fatty acid metabolism in bovine (Guo et al., 2017).

Transcriptomic studies have shown that the expression of many miRNAs is species specific and tissue specific, indicating that miRNAs may have potential roles in organ and tissue development, metabolism, immune response (Lawless et al., 2014), milk production traits, and fertility (Fatima and Morris, 2013). In beef cattle, differences in the expression pattern of miRNAs have been identified in animals with different amounts of subcutaneous fat, which could indicate a potential regulatory role of these molecules in the development of adipose tissue (Jin et al., 2010) and fat metabolism (Romao et al., 2014). These studies have identified numerous miRNAs expressed in cattle, but the miRNA regulatory mechanisms that underlie these phenotypes are unclear.

A recent integrative analysis of miRNA–mRNA co-expression in this Nelore population revealed several genes and miRNAs as candidate regulators of intramuscular fat deposition. Glucose metabolism and inflammation processes were the main pathways found to influence intramuscular fat deposition in Nelore beef cattle (Oliveira et al., 2018). Furthermore, previous RNAseq studies including this population identified differences in the skeletal muscle transcriptome profile associated with extreme values of fatty acid content. Oleic acid and CLA-c9t11 content had significant effects on the expression level of genes related to oxidative phosphorylation, cell growth, survival, and migration (Cesar et al., 2016). However, there are no studies about miRNA–mRNA co-expression related to fatty acid composition in bovines.

The integration of previous transcriptomic studies (Cesar et al., 2016) with miRNA expression data information may help provide a better understanding of the molecular mechanisms involved in the variation of FA content and deposition. Therefore, the goal of this study was to perform an integrative miRNA and mRNA expression analysis in skeletal muscle of beef cattle to unravel novel regulatory networks and signaling pathways involved in fatty acid biosynthesis and composition.

MATERIAL AND METHODS

Ethics Statement

Experimental procedures were carried out in accordance with the relevant guidelines provided by the Institutional Animal Care and Use Committee Guidelines of the Embrapa Pecuária Sudeste—Protocol CEUA 01/2013. The Ethical Committee of the Embrapa Pecuária Sudeste (São Carlos, São Paulo, Brazil) approved all experimental protocols (approval code CEUA 01/2013) prior to the conduction of the study.

Phenotypic Data

Description of phenotypic data and genomic heritability for oleic acid (OA (C18:1 cis9) and conjugated linoleic acid (CLA-c9t11) content from Nelore steers were previously reported (Cesar et al., 2014), with genomic heritability mean values of 0.16 ± 0.11 and 0.04 ± 0.09, respectively.

Animals used in the RNAseq and miRNAseq analyses were selected for extreme values of OA and CLA based on the rank of their genomic estimated breeding values (GEBVs) in a larger population of 386 Nellore steers. The GEBVs were calculated by GenSel software using the Bayes B approach (Cesar et al., 2014). A total of 30 animals with extreme GEBVs for OA and CLA content were selected separately: the top 13 high (H-OA) and 15 low (L-OA) animals for OA content, and the top 15 high (H-CLA) and 15 low (L-CLA) animals for CLA content. The Nelore steers used in this study are exactly the same group of animals used in Cesar et al. (2016).

mRNA Expression Data

The processing and analysis of mRNA expression data from skeletal muscle from the same population of animals used in this study were previously described in Cesar et al. (2016), with the sample accession of mRNA expression data of PRJEB13188. In total, 16,710 genes were identified as expressed in skeletal muscle for OA, and 16,530 genes were expressed in skeletal muscle for CLA. These genes were used for co-expression analysis.

miRNA Expression Data

The processing and analysis of miRNA expression data from skeletal muscle and miRNA target predictions used in this study followed the same procedures previously described in De Oliveira et al. (2018). Therefore, we will give a brief description of the analyses carried out.

In brief, sequencing of miRNA cDNA libraries was conducted on a MiSeq (Illumina, San Diego, CA) with MiSeq Reagent Kit 50 cycles in the Laboratory Multiuser ESALQ in Piracicaba/SP/Brazil, according to the protocol described by Illumina. The FastQC tools (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and FASTX (http://hannonlab.cshl.edu/fastx-toolkit) were used to check the quality of reads, and reads were subjected to alignment to bovine genome reference UMD version 3.1 (Ensembl 84: Mar 2016) through the software miRDeep2 version 2.0.0.7 (Friedländer et al., 2008). The reads were then mapped to regions of the genome using the Bowtie tool (Langmead et al., 2009), built into miRDeep2 software.

Differentially Expressed miRNAs

Differentially expressed (DE) miRNAs were identified for OA and CLA content phenotypes from a total of 30 small RNA libraries derived from skeletal muscle (N = 13 H-OA, N = 15 L-OA and N = 15 H-CLA, N = 15 L-CLA) using DESeq2 software (Love et al., 2014). The Benjamini–Hochberg method (Benjamini and Hochberg, 1995) was used to control for the rate of false positive (FDR; 10%) due to the number of genes and miRNAs tested. We set an FDR threshold of 0.1 (i.e., 10% of false positives are expected) to correct for false positives, avoiding to lose information, as these are exploratory analyses that should indicate biological responses to be in the future verified.

miRNA Target Predictions and Functional Enrichment Analysis

The target genes of DE miRNAs from skeletal muscle were predicted with TargetScan (Agarwal et al., 2015) and miRanda (Betel et al., 2010) software. In order to predict the potential regulatory target transcripts, the target genes were filtered by skeletal muscle (Cesar et al., 2016) mRNA expression data previously analyzed on the same set of samples. Functional enrichment analysis of target genes was performed by WebGestalt (Wang et al., 2017) using *Bos taurus* and the overrepresentation enrichment analysis (ORA) as organism and method of interest.

miRNA and mRNA Co-Expression Approaches: Weighted Gene Co-Expression Network Analysis (WGCNA) and Partial Correlation With Information Theory (PCIT)

WGCNA

Co-expression networks were constructed by WGCNA (Langfelder and Horvath, 2008) v1.36 package in RStudio environment using miRNA (N = 13 H-OA, N = 15 L-OA and N = 15 H-CLA, N = 15 L-CLA) and mRNA (N = 13 H-OA, N = 15 L-OA and N = 15 H-CLA, N = 15 L-CLA) skeletal muscle expression data.

miRNA and mRNA networks were constructed separately for high (H) and low (L) fatty acid groups. miRNA network construction and module detection used the step-by-step signed network construction with a soft threshold of $\beta = 6$ ($R^2 > 0.90$) and a minimum module size of 5. The same approach was adopted for mRNA signed network construction with a soft threshold $\beta = 6$ ($R^2 > 0.91$) and a minimum module size of 30. Five was chosen as the minimum module size for the miRNAs due to the smaller size of the miRNA transcriptome relative to the mRNA transcriptome (Langfelder and Horvath, 2008; Betel et al., 2010). The topological overlap distance calculated from the adjacency matrix is then clustered with the average linkage hierarchical clustering. The default minimum cluster merge height of 0.25 was retained.

An integrative analysis was performed, in which miRNA module eigengenes (MEs) and mRNA MEs for high and low fatty acid groups were correlated with one another by calculating the Pearson correlations. miRNA and mRNA modules with a negative correlation and a p-value < 0.10 were selected for functional enrichment analysis. miRNA modules that were significantly correlated were then further explored to identify hub miRNAs. Hub miRNAs were selected based on the top five greatest module membership (MM) values.

In order to better understand the biological significance of the modules identified, the functional enrichment analysis of genes and miRNA target genes were performed by WebGestalt (Wang et al., 2017) web tool. The functional enrichment analysis used the list of target genes from hub miRNAs selected from miRNA modules that were negatively correlated with mRNA modules. Co-expression networks among hub miRNAs and the GO terms of the target genes were constructed in Cytoscape v.3.3.0 (Cline et al., 2007).

PCIT: Differential Hubbing (DH), Regulatory Impact Factor (RIF), and Phenotypic Impact Factor (PIF) Metrics

The gene list used for PCIT analyses included all miRNAs and mRNAs detected in our study, but only those with a direct and partial correlation greater than 0.90 were used for the differential hubbing (DH) analyses. The DH was computed by the difference of significant connections of mRNA and miRNA between the high and low OA and CLA groups. Functional enrichment analysis was performed on the top five differential hubbing genes, or for miRNAs based on the list of predicted target genes, to determine if these specific genes/miRNAs have biological relevance to fatty acid composition.

The regulatory impact factor (RIF) and phenotypic impact factor (PIF) scores were calculated as described in Reverter et al. (2010) to predict which transcripts were potential regulators of gene expression differences between the high and low OA and CLA groups. Regulatory impact factor 1 (RIF1) ranks genes as potential regulators of networks based largely on changes in correlations between two different states, while regulatory impact factor 2 (RIF2) ranks genes with more emphasis on how expression level changes between two different states (Reverter et al., 2010). Phenotypic impact factor (PIF) values were used to identify and rank genes based on the magnitude of gene expression and the difference in the expression of that gene between two treatments (Reverter et al., 2010). The RIF calculations presented here were modified from the original method, and the complete list of expressed mRNA or miRNA was tested as potential regulators, and only mRNAs or miRNAs with a significant partial correlation of 0.90 from PCIT were included in the RIF and PIF score estimates; as described in Cesar et al. (2016). PIF score estimates were ranked to select the top 10 PIF regulators of all genes in the dataset.

RESULTS

Phenotypic and miRNA Expression Data

Genomic estimated breeding values (GEBVs) and the number of normalized mapped miRNA reads for Nelore steers genetically divergent for two FA content, oleic acid (C18:1 cis9) and conjugated linoleic acid (CLA-c9t11), are shown in **Table 1**. Sample sizes of extreme Nelore steers used for OA and CLA analyses were slightly different due to data availability. A Student's t-test was previously performed by Cesar et al. (2016) to evaluate

TABLE 1 | Genomic estimated breeding values (GEBVs) and number of normalized mapped miRNA reads for Nelore steers with divergent fatty acid content groups.

Groups[1]	GEBVs	miRNA mapped reads	Groups[2]	GEBVs	miRNA mapped reads
OA (C18:1 cis9)					
H-OA1	2.11	970,350	L-OA1	–7.06	605,324
H-OA2	1.54	745,742	L-OA2	–4.98	1,308,497
H-OA3	3.20	1,009,408	L-OA3	–5.51	703,560
H-OA4	3.25	1,051,535	L-OA4	–2.41	525,015
H-OA5	2.16	1,052,657	L-OA5	–1.54	594,924
H-OA6	2.08	1,289,704	L-OA6	–2.95	563,260
H-OA7	1.58	610,678	L-OA7	–2.43	349,772
H-OA8	1.92	564,255	L-OA8	–7.03	557,486
H-OA9	4.14	555,136	L-OA9	–3.48	656,226
H-OA10	2.61	607,145	L-OA10	–2.67	685,506
H-OA11	3.89	1,035,882	L-OA11	–1.63	690,291
H-OA12	2.84	520,790	L-OA12	–8.07	748,090
H-OA13	1.76	584,448	L-OA13	–4.14	689,313
H-OA14	*	*	L-OA14	–7.36	602,604
H-OA15	*	*	L-OA15	–2.91	732,483
Mean	**2.16**	**745,742**	**Mean**	**–3.48**	**656,226**
CLA (CLA-c9t11)					
H-CLA1	0.008	439,844	L-CLA1	–0.008	457,706
H-CLA2	0.012	1,152,736	L-CLA2	–0.013	532,780
H-CLA3	0.008	938,908	L-CLA3	–0.017	988,632
H-CLA4	0.012	1308,497	L-CLA4	–0.011	521,652
H-CLA5	0.011	235,852	L-CLA5	–0.010	300,196
H-CLA6	0.016	874,538	L-CLA6	–0.009	1,264,789
H-CLA7	0.022	976,093	L-CLA7	–0.016	685,506
H-CLA8	0.025	1,039,805	L-CLA8	–0.013	702,016
H-CLA9	0.009	555,136	L-CLA9	–0.015	605,442
H-CLA10	0.009	918,555	L-CLA10	–0.013	677,737
H-CLA11	0.012	675,360	L-CLA11	–0.009	522,944
H-CLA12	0.012	772,395	L-CLA12	–0.010	658,923
H-CLA13	0.019	713,501	L-CLA13	–0.010	752,121
H-CLA14	0.019	500,697	L-CLA14	–0.018	1,306,312
H-CLA15	0.014	636,795	L-CLA15	–0.010	511,704
Mean	**0.012**	**772,395**	**Mean**	**–0.011**	**658,923**

[1]*Nelore samples of high (H) oleic acid (OA) and conjugated linoleic acid (CLA) groups.*
[2]*Nelore samples of low (L) oleic acid (OA) and conjugated linoleic acid (CLA) groups.*
**Data not available.*

the mean differences between the high and low FA groups, and significant differences (p-value < 0.05) were observed for both OA and CLA content.

On average, 85% of miRNA reads for OA and CLA were mapped to the *B. taurus* UMD 3.1 genome assembly (Ensembl 84: Mar 2016). In total, 404 OA and 386 CLA mature miRNAs were detected by miRDeep2 software (**Supplementary Table 1**).

Differentially Expressed miRNAs and Target Gene Identification

We identified 137 and 131 unique mature miRNA sequences with non-zero expression levels according to DESeq criteria for samples in OA (**Supplementary Table 2**) and CLA (**Supplementary Table 3**) groups, respectively, which were used in differential expression and co-expression analyses. The miRNAs bta-miR-126-5p and bta-miR-2419-5p were upregulated in the L-OA and H-CLA groups, respectively (**Table 2**).

Among the 3,619 bta-miR-126-5p target genes (**Supplementary Table 4**) identified in the bovine genome, 162 were previously identified as DE in this same population (Cesar et al., 2016). *SCD* (stearoyl-CoA desaturase), *CDS2* (CDP-diacylglycerol synthase 2), *FAR2* (fatty acyl CoA reductase 2), and *NAB1* (NGFI-A binding protein 1) were included in this list of DE genes, well known to be related to biological processes associated with fatty acid composition. Regarding the 365 bta-miR-2419-5p target genes (**Supplementary Table 4**), 16 were also previously identified as DE (Cesar et al., 2016). Among these genes are included *CAV3* (caveolin 3), *JMJD1C* (jumonji domain containing 1C), *FOXO6* (forkhead box O6), and *PRKAG2* (protein kinase AMP-activated non-catalytic subunit gamma 2). These bta-miR-126-5p DE target genes were also enriched for genes involved in biological processes associated with fatty acid composition.

miRNA and mRNA Co-Expression

Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA was used to identify potential regulatory networks related to OA and CLA content in skeletal muscle. To this end, miRNA and mRNA expression data from animals with extreme

TABLE 2 | Differentially expressed miRNAs identified by miRDeep2 for Nelore steers with divergent fatty acid content groups and the number of predicted target genes.

miRNA	log2FC[1]	padj[2]	High[3]	Low[4]	Predicted target genes[5]
OA (C18:1 cis9)					
bta-miR-126-5p	0.4346	0.0987	3,334.235	4,682.251	3,619
CLA (CLA-c9t11)					
bta-miR-2419-5p	–1.18	0.0041	16,108.8	4,182.543	365

[1]log2 fold change.
[2]False discovery rate adjusted p-values by Benjamini–Hochberg (1995) methodology.
[3,4]Normalized miRNA mean counts of high and low fatty acid content groups.
[5]Number of predicted target genes by TargetsSan and miRanda software.

GEBVs of OA and CLA were evaluated separately. After quality control, a total of 137 miRNAs and 16,710 mRNAs were analyzed for OA network construction, while a total of 131 miRNAs and 16,530 mRNAs were used for CLA network construction.

A total of 52 mRNA modules were identified in the H-OA group (**Figure S1A**), while in the L-OA group, 95 mRNA modules were identified (**Figure S1B**). Regarding miRNA, 12 and 9 modules were identified, respectively, in the H-OA (**Figure S2A**) and L-OA (**Figure S2B**) groups. For CLA, 52 and 28 mRNA modules in the H-CLA (**Figure S3A**) and L-CLA (**Figure S3B**) groups were identified, respectively. CLA miRNA network analysis identified six and five miRNA modules in the H-CLA (**Figure S4A**) and L-CLA (**Figure S4B**), respectively.

In order to investigate miRNA–mRNA interactions, mRNA module eigengenes (MEs), which represent the sum of gene expression profiles of each module, were correlated with miRNA MEs. The mRNA and miRNA modules with negative correlations and a nominal p-value < 0.10 were selected for further investigation. The focus on negative correlation between mRNA and miRNA modules was based on the fact that the predominant and canonical effect of miRNAs on gene expression is through mRNA downregulation, which would equate to a negative miRNA–mRNA expression correlation (Guo et al., 2010). In the H-OA group, four negative correlations between mRNA and miRNA MEs were identified, while in the L-OA group, eight negative correlations were observed (**Table 3**). In the H-CLA group, seven negative correlations between miRNA and mRNA MEs were observed, while in the L-CLA group, three negative correlations were observed (**Table 3**).

Hub miRNAs are the miRNAs with higher connectivity inside the module and are probably more informative (Filteau et al., 2013). In order to find hub miRNAs involved in the co-expression networks, we selected the top five miRNAs representing the greatest module membership (MM) values from miRNA

TABLE 3 | Signaling pathways of miRNA module eigengenes (MEs) negatively correlated with mRNA MEs for Nelore steers with divergent fatty acid content groups.

Group	mRNA MEs	Corr	p-value[1]	miRNA MEs	Signaling pathways[2]	p-value[1]
H-OA	paletvioletred3	–0.7	0.01	Magenta	Insulin resistance	6.23e–06
	grey	–0.6	0.04	Black	MAPK signaling pathway	5.54e–07
	lightsteelblue1	–0.8	0.00	Turquoise	Insulin resistance	1.64e–04
	bisque4	–0.7	0.00			
L-OA	coral1	–0.7	0.00	Turquoise	AMPK signaling pathway	7.11e–07
	mediumpurple2	–0.6	0.01			
	darkolivegreen	–0.7	0.00	Pink	Insulin signaling pathway	9.76e–08
	orangered	–0.5	0.04			
	lightcyan1	–0.5	0.05	Blue	Proteoglycans in cancer	1.17e–04
	mediumorchid	–0.6	0.01			
	lightcyan1	–0.6	0.03	Yellow	Wnt signaling pathway	1.84e–05
	lightcyan1	–0.7	0.00	Green	Insulin signaling pathway	2.89e–04
H-CLA	cyan	–0.6	0.03	Turquoise	ns	*
	brown	–0.6	0.02	Red	Insulin resistance	7.9e–05
	plum2	–0.6	0.03	Blue	Insulin resistance	1.1e–05
	turquoise	–0.6	0.03			
	darkred	–0.7	0.00	Brown	ns	*
	steelblue	–0.5	0.04			
	thistle1	–0.6	0.02	Black	Insulin signaling pathway	2.73e–07
L-CLA	purple	–0.5	0.04	Blue	Insulin signaling pathway	6.68e–06
	white	–0.5	0.04			
	orange	–0.5	0.04	Red	MAPK signaling pathway	1.71e–04

[1]Nominal p-value.
[2]Signaling pathways of target genes from hub miRNAs selected based on greatest module membership values.
**NS, non-significant.*

modules negatively correlated with mRNA modules. **Table 3** shows signaling pathways obtained from WebGestalt software based on enrichment of the genes from miRNA modules and the target genes for hub miRNAs.

In the H-OA group, genes from mRNA modules were not significantly enriched for any Gene Ontology terms, while miRNA target genes from magenta and turquoise modules were significantly enriched for insulin resistance (**Supplementary Table 5**). miRNA target genes from black module were enriched for the MAPK signaling pathway (**Supplementary Table 5**). In the L-OA group, genes from the dark olive green module were significantly enriched for fatty acid degradation processes, while miRNA target genes from turquoise, pink, and blue modules were enriched for AMPK, insulin signaling pathway, and proteoglycans in cancer, respectively (**Supplementary Table 5**).

In the H-CLA group, genes from the mRNA modules brown and plum were significantly enriched for insulin resistance and steroid biosynthesis, respectively, while miRNA target genes from red, blue, and black modules were enriched for insulin resistance and insulin signaling pathway (**Supplementary Table 6**). In the L-CLA group, miRNA target genes from blue and red modules were enriched for insulin and MAPK signaling pathway, respectively (**Supplementary Table 6**).

Figures 1 and **2** show co-expression networks for hub miRNAs enriched for signaling pathways related to FA composition in H-OA (**Figure 1**) and L-OA groups (**Figure 2**) and in H-CLA (**Figure 3**) and L-CLA (**Figure 4**) groups in Nelore cattle.

Partial Correlation With Information Theory (PCIT)

PCIT, regulatory impact factor (RIF), and phenotypic impact factor (PIF) were used to score genes as potential regulators of signaling pathways between high and low FA groups. Differential hubbing (DH) represents the number of significant partial correlations that a gene has between two phenotypic states (Hudson et al., 2009). DH values of all genes and miRNAs used in PCIT analysis are in **Supplementary Table 7**. **Table 4** shows the top five negative and positive extreme hubbing genes when comparing high and low OA and CLA groups. Functional enrichment analysis was performed on the top five DH genes to determine if they have biological relevance to fatty acid composition. The top 10 negatively and positively hubbed genes for OA and CLA content and their associated GO terms are shown in **Supplemental Table 8**.

The bta-mir-339a and *FRAT1* are among the top negatively hubbed genes for OA, associated with glycogen synthase kinase-3 binding protein (IPR008014) and Wnt signaling pathway (bta04310). *GAL3ST3* and *ATP6V0E1* are among the top positively hubbed genes, associated with the glycolipid biosynthetic process (GO: 0009247) and generation of metabolites and energy, respectively. *KAT5* is among the top negatively hubbed genes for CLA, associated with proteasome-mediated ubiquitin-dependent protein catabolic process (GO: 0043161), while *TMEM115* is associated with protein glycosylation (GO: 0006486). *PSMG1* is among the top positively hubbed genes and associated with proteasome assembly (GO: 0043248).

For OA, a total of 14,900 transcripts had negative RIF1 values, whereas 1,948 transcripts had positive values. For RIF2, 16,684 transcripts had negative values, while 127 transcripts had positive values (**Supplementary Table 9**). For CLA, a total of 14,171 transcripts had negative RIF1 values, while 1,891 transcripts had positive values. For RIF2, 108 transcripts had negative values, while 16,512 transcripts had positive values (**Supplementary Table 10**).

Table 5 shows the top five negative and positive genes identified by RIF1 and RIF2 score by contrasting H and L groups for OA content. The RIF1 analysis identified putative regulators for OA content GO terms associated with muscling, e.g., muscle contraction and actin filament organization (*TPM1*, *TPM2*, and *MYL1*), and with fatness, e.g., glycolytic process and ATP biosynthetic process (*ALDOA*). The RIF2 analysis identified bta-mir-10b as a negative putative regulator, as well as the genes *ACTN2* and *TNNT1*, associated with skeletal muscle contraction GO terms. The same group of genes identified as positive RIF1 regulators is among the top positive RIF2 regulators. Functional enrichment analysis was performed using DAVID software on the top five genes to determine their biological relevance on fatty acid composition (**Supplementary Table 9**).

Table 6 shows the top negative and positive RIF1 and RIF2 genes when high and low groups for CLA content were

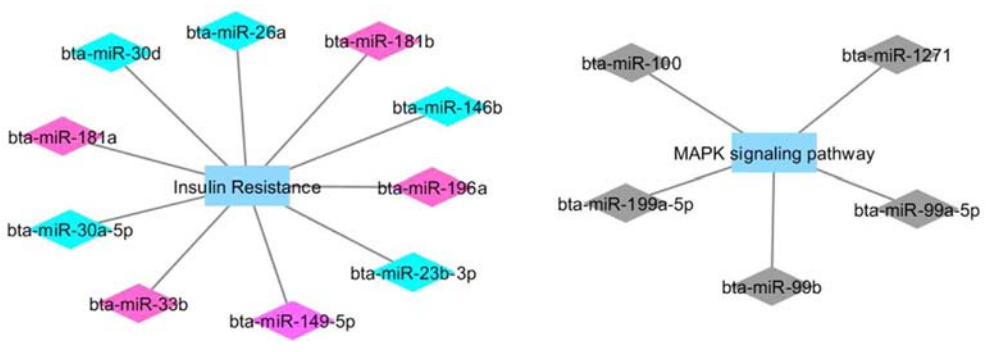


FIGURE 1 | Co-expression networks of H-OA group from Nelore cattle. Colored diamonds represent the top five hub miRNAs within each module, and colored rectangles represent the signaling pathways associated (p-value ≤ 0.10) with hub miRNA target genes.

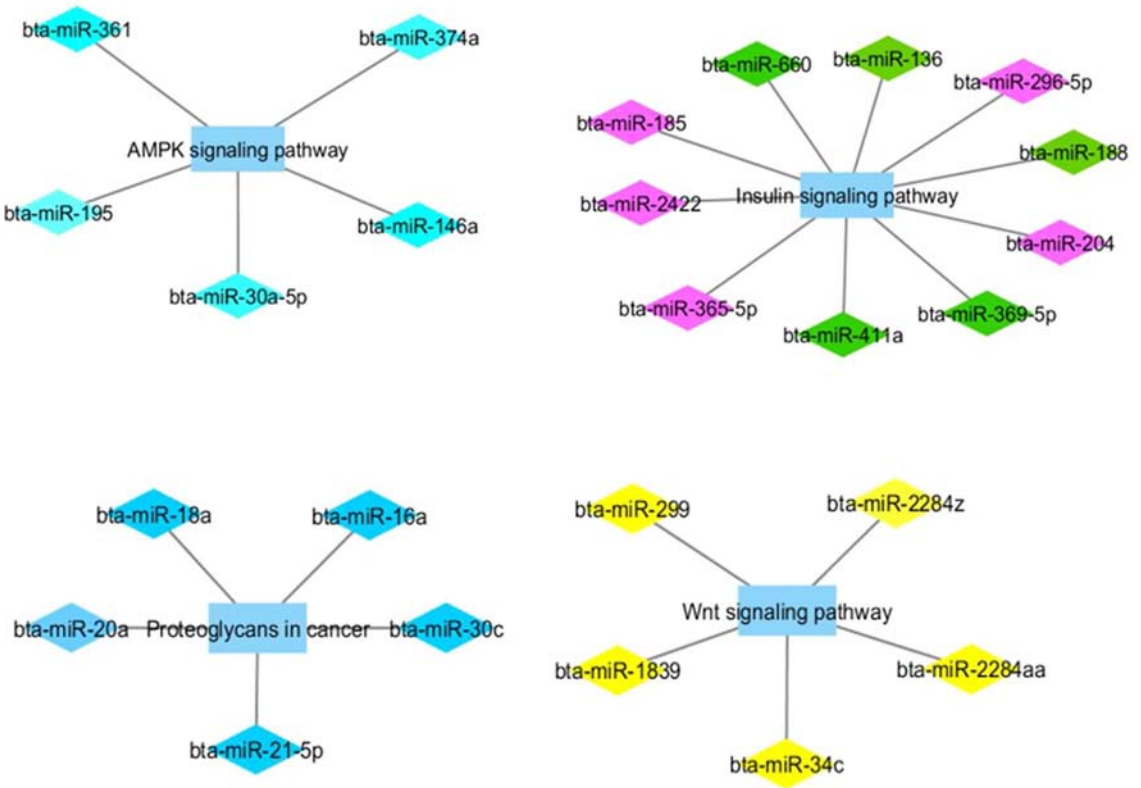


FIGURE 2 | Co-expression networks of L-OA group from Nelore cattle. Colored diamonds represent the top five hub miRNAs within each module, and colored rectangles represent the signaling pathways associated (p-value ≤ 0.10) with hub miRNA target genes.

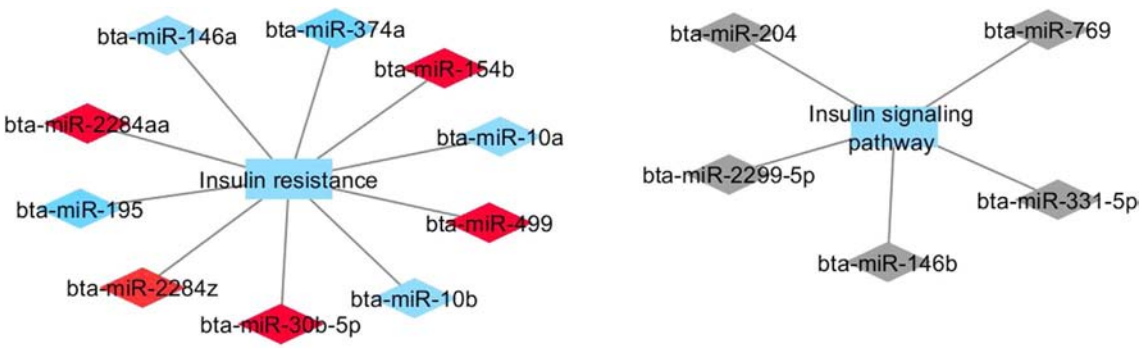


FIGURE 3 | Co-expression networks of H-CLA group from Nelore cattle. Colored diamonds represent the top five hub miRNAs within each module, and colored rectangles represent the signaling pathways associated (p-value ≤ 0.10) with hub miRNA target genes.

contrasted. RIF1 analysis identified putative negative regulators for CLA content genes related to regulation of gene expression (*GNRH1*) and muscling, e.g., actin filament organization (*TRPV4*). The genes *ALDOA*, *CKM*, and *TPM* were identified as positive RIF1 regulators and were also identified as negative RIF2 regulators. RIF2 analysis identified positive putative regulator genes enriched for GO terms associated with muscling (*MYH1*, *DES*, and *PDLIM3*). Functional enrichment analysis on the top five genes is presented in **Supplementary Table 10**.

For OA and CLA, PIF analysis identified 13,480 and 15,531 transcripts, respectively (adjusted p-value < 0.05), which could significantly impact fatty acid composition. **Table 7** lists the top 10 regulators ranked by PIF analysis for OA and CLA content. Functional enrichment analysis to determine the biological relevance on fatty acid composition of the top 10 genes is presented in **Supplementary Table 11**.

Interestingly, among the top 10 regulators identified by PIF analyses for OA and CLA content is the same group of genes

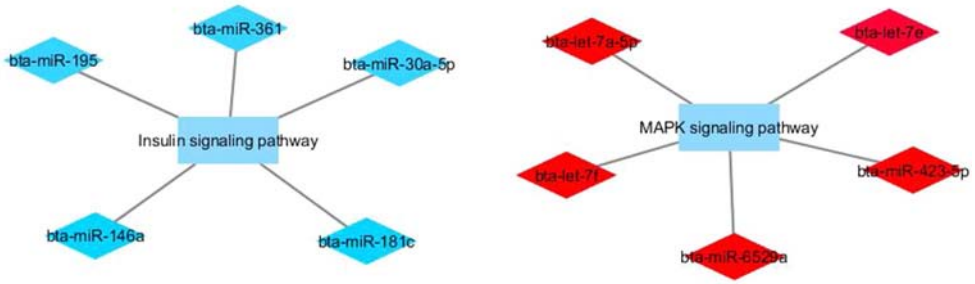


FIGURE 4 | Co-expression networks of L-CLA group from Nelore cattle. Colored diamonds represent the top five hub miRNAs within each module, and colored rectangles represent the signaling pathways associated (p-value ≤ 0.10) with hub miRNA target genes.

TABLE 4 | Top negative and positive extreme hubbing genes groups for oleic acid (OA) and conjugated linoleic acid (CLA) content.

Ensembl gene ID	Gene symbol	Description	DH*
Negative hub genes OA			
ENSBTAG00000040485	*THEG*	Theg spermatid protein	–905
ENSBTAG00000031806	*ZNF48*	Zinc finger protein 48	–826
ENSBTAG00000007692	–	–	–799
ENSBTAG00000044830	bta-mir-339a	bta-mir-339a	–770
ENSBTAG00000007956	*FRAT1*	Frequently rearranged advanced T-cell lymphomas 1	–753
Positive hub genes OA			
ENSBTAG00000015100	*ATP6V0E1*	ATPase H+ transporting V0 subunit e1	2946
ENSBTAG00000025221	TEX30	Testis expressed 30	2,925
ENSBTAG00000014337	*EIF2S3*	Eukaryotic translation initiation factor 2 subunit 3	2,911
ENSBTAG00000044066	*C1orf101*	Chromosome 1 open reading frame	2,903
ENSBTAG00000002772	*GAL3ST3*	Galactose-3-*O*-sulfotransferase 3	2,902
Negative hub genes CLA			
ENSBTAG00000005578	*KAT5*	Lysine acetyltransferase 5	–2,126
ENSBTAG00000046307	*CEBPD*	CCAAT enhancer binding protein delta	–2,112
ENSBTAG00000003849	*GNL2*	G protein nucleolar 2	–2,060
ENSBTAG00000045822	–	–	–1,997
ENSBTAG00000002634	*TMEM115*	Transmembrane protein 115	–1,987
Positive hub genes CLA			
ENSBTAG00000013600	*PSMG1*	Proteasome assembly chaperone	1,848
ENSBTAG00000015042	*WDR55*	WD repeat domain	1,809
ENSBTAG00000032684	*TCTEX1D2*	Tctex1 domain containing 2	1,794
ENSBTAG00000006321	*RTCA*	RNA 3′-terminal phosphate	1,740
ENSBTAG00000045544	*LOC107131189*	Eukaryotic translation initiation factor	1,692

**DH, differential hubbing.*

(*ACTA1*, *TTN*, *MYH1*, *ALDOA*, *CKM*, and *NEB*) with Gene Ontology (GO) terms associated with muscling (*ACTA1*, *MYH*, and *MYLPF*) and fatness and fiber type (*ALDOA*).

DISCUSSION

The purpose of this study was to investigate the complex interactions between miRNAs and mRNAs in bovine skeletal muscle associated with variation in oleic acid (OA) and conjugated linoleic acid (CLA-c9t11) content. This was accomplished by performing differential miRNA expression analysis and two different gene co-expression approaches, weighted gene co-expression network analysis (WGCNA) and partial correlation with information theory (PCIT). Furthermore, we sought to identify the key drivers of gene expression networks. Analyses were focused on these two fatty acids due to their importance in many biological processes and beneficial effects on metabolic diseases and human health (Laaksonen et al., 2005b). Also, a significant number of differentially expressed genes in response to OA (1134) and CLA (872) content were identified in this same population previously (Cesar et al., 2016). OA is a monounsaturated fatty acid present in membrane phospholipids, triglycerides, and cholesterol and is associated with protection against heart disease (Laaksonen et al., 2005b), while CLA antidiabetic effects are mediated *via* anti-inflammatory processes in white adipose tissue (Moloney et al., 2007).

Integration of miRNA and mRNA co-expression from next-generation sequencing data from the same individuals revealed the possible regulatory roles of miRNAs on fatty acid

TABLE 5 | Top negative and positive genes identified by regulatory impact factor 1 (RIF1) and regulatory impact factor 2 (RIF2) score for oleic acid (OA) content.

Ensembl gene ID	Gene symbol	Description	RIF1 *Z*-score*
Top negative RIF1			
ENSBTAG00000000020	*TRPV3*	Transient receptor potential cation channel subfamily V member 3	–0.046
ENSBTAG00000000030	*RDM1*	RAD52 motif containing 1	–0.046
ENSBTAG00000000102	*GPR75*	G protein-coupled receptor 75	–0.046
ENSBTAG00000000008	*KCNJ1*	Potassium voltage-gated channel subfamily J member 1	–0.046
ENSBTAG00000000168	–	–	–0.046
Top positive RIF1			
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	94.532
ENSBTAG00000011424	*TPM2*	Tropomyosin 2 (beta)	69.720
ENSBTAG00000013921	*CKM*	Creatine kinase	27.167
ENSBTAG00000009707	*MYL1*	Myosin light chain 1	23.357
ENSBTAG00000005373	*TPM1*	Tropomyosin 1 (alpha)	22.028
Top negative RIF2			
bta-miR-10b	bta-mir-10b	bta-mir-10b	–1.377
ENSBTAG00000007782	*MYOT*	Myotilin	–0.238
ENSBTAG00000009696	*ACTN2*	Alpha-actinin-2	–0.074
ENSBTAG00000000678	*CSDE1*	Cold shock domain containing E1	–0.060
ENSBTAG00000006419	*TNNT1*	Troponin T	–0.050
Top positive RIF2			
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	106.069
ENSBTAG00000011424	*TPM2*	Tropomyosin 2 (beta)	54.632
ENSBTAG00000013921	*CKM*	Creatine kinase	28.422
ENSBTAG00000005373	*TPM1*	Tropomyosin 1 (alpha)	22.117
ENSBTAG00000009707	*MYL1*	Myosin light chain 1	16.740

**RIF1 and RF2 scores are presented as Z-score-normalized values.*

TABLE 6 | Top negative and positive genes identified by regulatory impact factor 1 (RIF1) and regulatory impact factor 1 (RIF2) score for conjugated linoleic acid (CLA) content.

Ensembl gene ID	Gene symbol	Description	RIF1 *Z*-score*
Top negative RIF1			
ENSBTAG00000000164	*GNRH1*	Gonadotropin releasing hormone 1	–0.0633
ENSBTAG00000000582	*LY6G6E*	Lymphocyte antigen 6 complex	–0.0633
ENSBTAG00000000982	*C22H3orf84*	Chromosome 22 c3orf84 homolog	–0.0633
ENSBTAG00000000031	*TRPV4*	Transient receptor potential cation channel subfamily V member 4	–0.0633
ENSBTAG00000000535	*PCNX2*	Pecanex homolog 2	–0.0633
Top positive RIF1			
ENSBTAG00000026986	*TTN*	Titin	25.8861
ENSBTAG00000006907	*NEB*	Nebulin	29.2850
ENSBTAG00000011424	*TPM2*	Tropomyosin 2 (beta)	36.2570
ENSBTAG00000013921	*CKM*	Creatine kinase	36.8083
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	89.4190
Top negative RIF2			
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	–108.25
ENSBTAG00000013921	*CKM*	Creatine kinase	–38.19
ENSBTAG00000011424	*TPM2*	Tropomyosin 2 (beta)	–38.08
ENSBTAG00000005373	*TPM1*	Tropomyosin 1 (beta)	–24.17
ENSBTAG00000046332	*ACTA1*	Actin alpha 1	–20.98
Top positive RIF2			
ENSBTAG00000018204	*MYH1*	Myosin heavy chain 1	11.5144
ENSBTAG00000005353	*DES*	Desmin	6.2929
ENSBTAG00000006907	*NEB*	Nebulin	4.5481
ENSBTAG00000006823	*CMYA5*	Cardiomyopathy associated 5	1.9259
ENSBTAG00000017183	*PDLIM3*	PDZ and LIM domain 3	1.3030

**RIF1 and RIF2 scores are presented as Z-score-normalized values.*

TABLE 7 | Top 10 regulators identified by phenotypic impact factor (PIF) analysis with FDR adjusted *p*-value for oleic acid (OA) and conjugated linoleic acid (CLA) content.

Ensembl gene ID	Gene symbol	Description	PIF	adj.*p*.Val
OA				
bta-miR-10b	bta-miR-10b	bta-miR-10b	66,197	1.10e–01*
ENSBTAG00000046332	*ACTA1*	Actin, alpha 1, skeletal muscle	20,402	1.27e–005
ENSBTAG00000026986	*TTN*	Titin	8,820	4.81e–007
ENSBTAG00000018204	*MYH1*	Myosin, heavy chain 1, skeletal muscle, adult	7,401	8.64e–008
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	7,199	9.68e–006
ENSBTAG00000011424	–	–	4,547	2.42e–005
ENSBTAG00000013921	*CKM*	Creatine kinase, M-type	3,610	5.41e–006
bta-miR-486	bta-miR-486	bta-mir-486	3,506	4.27e–001*
ENSBTAG00000006907	*NEB*	Nebulin	2,864	1.91e–007
ENSBTAG00000021218	*MYLPF*	Myosin light chain, phosphorylatable, fast skeletal muscle	2,111	6.31e–005
CLA				
bta-miR-10b	bta-miR-10b	bta-miR-10b	92,224	9.53e–003
ENSBTAG00000046332	*ACTA1*	Actin, alpha 1, skeletal muscle	39,297	1.48e–009
ENSBTAG00000012927	*ALDOA*	Aldolase, fructose-bisphosphate A	21,787	7.32e–004
ENSBTAG00000026986	*TTN*	Titin	18,487	2.16e–007
ENSBTAG00000018204	*MYH1*	Myosin, heavy chain 1, skeletal muscle, adult	14,134	1.65e–011
ENSBTAG00000011424	–	–	9,845	2.65e–004
ENSBTAG00000013921	*CKM*	Creatine kinase, M-type	7,206	1.41e–006
ENSBTAG00000006907	*NEB*	Nebulin	5,455	6.68e–007
ENSBTAG00000021218	*MYLPF*	Myosin light chain, phosphorylatable, fast skeletal muscle	4,229	1.27e–004
bta-mir-486	bta-mir-486	bta-mir-486	3,777	1.18e–003

**Non-significant adj p-values.*

composition. WGCNA allowed the identification of miRNA and mRNA modules that were negatively correlated with each other, which may indicate that FA composition is modulated by specific miRNA–mRNA interactions. However, only one mRNA module in the L-OA group and two mRNA modules in the H-CLA group presented functional enrichment for fatty acid degradation, insulin resistance, and steroid biosynthesis. Therefore, the subsequent analyses were focused on the identification of key miRNAs that may be involved in co-expression networks and thereby in the regulation of fatty acid composition.

From differential expression analysis, the *SCD* gene, identified as a putative target gene of bta-miR-126-5p, was found to be upregulated in the H-OA content group in a previous skeletal muscle RNAseq study (Cesar et al., 2016). *SCD* is as a key enzyme in *de novo* lipogenesis (Scaglia and Igal, 2005), and its upregulation has been associated with deposition of unsaturated FAs. Thus, the upregulated bta-miR-126-5 in the L-OA group observed herein may explain the reduced *SCD* mRNA levels observed before (Cesar et al., 2016). These results suggest that *SCD* expression level is related to OA content and may suggest a new role for bta-miR-126-5p in the regulation of *SCD*. It is relevant to note that only negative miRNA–mRNA correlations were considered herein. However, it is important to emphasize that this was an exploratory study and should be complemented by other *in vitro* and *in vivo* analyses to better discriminate the mechanisms of miRNA regulation of gene expression.

Functional enrichment analysis indicated a relationship between insulin, insulin resistance, adipocytokine signaling pathway, and non-alcoholic fatty liver disease for *PRKAG2* gene, a target gene of bta-miR-2419-5p. The major effects of insulin on muscle and adipose tissue are related to carbohydrate, lipid, and protein metabolism (Dimitriadis et al., 2011). Insulin can decrease the rate of lipolysis, stimulate fatty acid synthesis, increase the uptake of triglycerides from blood into muscle and adipose tissue, and decrease the rate of fatty acid oxidation in muscle and liver (Dimitriadis et al., 2011). Metabolic diseases such as obesity and coronary disorders can be the consequence of insulin resistance, i.e., the inability of insulin to drive glucose into muscle and other tissues, which can be caused by excessive body fat deposition (Dimitriadis et al., 2011). Assuming that bta-miR-2419-5p expression level regulates CLA content and that this miRNA can regulate *PRKAG2* expression, it is possible that bta-miR-2419-5p can regulate insulin expression to ultimately respond to CLA content. The transcription factors (TFs) *FOXO1*, *FOXO3*, and *FOXO4* are also targets of bta-miR-2419-5p. Therefore, they may also be regulated by CLA content. *FOXO1* expression has been previously associated with CLA content in this population (Cesar et al., 2016).

Co-expression network analysis integrating miRNA and mRNA expression data revealed two miRNA modules (magenta and turquoise) in the H-OA group whose potential target genes were significantly enriched for the GO term insulin resistance. As discussed previously, it is known that insulin resistance can be caused by excessive fat deposition (Ortega et al., 2013). The miRNA magenta module grouped both bta-miR-181a and bta-miR-33a/b. The bta-miR-181a is from the same family

as bta-miR-181b, which has been reported to regulate the biosynthesis of bovine milk by targeting *ACSL1*, an important enzyme of milk lipid synthesis (Lian et al., 2016). Furthermore, bta-miR-33a/b has been reported to contribute to the regulation of fatty acid metabolism and the insulin signaling pathway (Dávalos et al., 2011), which indicates that the insulin signaling pathway may be impacted in the H-OA group.

The miRNA turquoise module contained both bta-miR-146b and bta-miR-26b. In a recent integrative analysis of miRNA–mRNA expression related to intramuscular fat (IMF) deposition in animals from this population, the bta-miR-146b was found to be downregulated in individuals with high IMF content, while the bta-miR-26b was identified by PCIT as a candidate regulatory gene that negatively regulates IMF deposition (Oliveira et al., 2018). IMF represents the amount of fat accumulated between muscle fibers or within muscle cells, and it is a determinant factor that affects meat quality (Wood et al., 2008). Despite the fact that in the present study IMF deposition between high and low groups was not statistically different (Cesar et al., 2016), we could not ignore the fact that OA, CLA, and other FAs comprise IMF. It has been reported that as intramuscular lipid content accumulates, there is a concomitant elevation in the concentration of oleic acid (Smith et al., 2006). Other hub miRNAs such as bta-miR-196a and bta-miR-30f (from the same family of hubs bta-miR-30d and bta-miR-30a-5p) have been reported to have a higher expression level in cattle that have higher amounts of IMF (Guo et al., 2017). Therefore, taking together the results from these and previous studies, we can highlight the role of bta-miR-146b, bta-miR-26b, bta-miR-30d, and bta-miR-196a in regulating muscle fatty acid composition.

Lipid accumulation can activate the immune system and inflammatory pathways due to the secretion of proinflammatory molecules by adipocytes (Lau, 2005). The mitogen-activated protein kinase (MAPK) cascade, enriched from the miRNA black module, is highly conserved and involved in various cellular functions, including responses to proinflammatory stimuli (Soares-Silva et al., 2016). Networks and DE genes related to immune system and inflammatory response were previously associated with high amounts of IMF in this population (Cesar et al., 2015; Oliveira et al., 2018). Among the hub miRNAs from the black module was the bta-miR100, which was previously found to be downregulated in animals with high IMF (Oliveira et al., 2018).

In the L-OA group, the target genes of two miRNA modules (pink and green) were enriched for GO terms associated with the insulin signaling pathway. Insulin is a hormone with a direct effect on lipid metabolism (Dimitriadis et al., 2011), which could explain why modules present in the high and low fatty acid groups are associated with insulin-related terms. However, we can observe an overrepresentation of the insulin signaling pathway in the low fatty acid group, while in the high fatty acid group, we can observe an overrepresentation of the insulin resistance pathway. As discussed before, insulin resistance has been linked to excessive body fat deposition and obesity (Dimitriadis et al., 2011), supporting our findings.

Hub miRNAs from the pink module (bta-miR-204 and bta-miR-365-5p) and from the green module (bta-miR-660) have been previously associated with adipose tissue in cattle (Gu et al., 2007). Other hub miRNAs from the green module (bta-miR-411a and bta-miR-136) were expressed at a higher level in Wagyu compared with Holstein cattle (Guo et al., 2017). Wagyu cattle accumulate large amounts of marbling and specifically monounsaturated fatty acids, of which oleic acid is primarily responsible for the soft fat (Smith et al., 2009). Taken together, these results may indicate a possible role of these miRNAs in post-transcriptional regulation of OA content.

Target genes of the miRNA turquoise module were enriched for GO terms associated with the AMPK signaling pathway. *AMPK* is a basic regulator of cellular and body energy metabolism and may enhance activity of mitochondrial proteins involved in oxidative metabolism (Thomson et al., 2008). Cesar et al. (2016) reported that several canonical pathways of oxidative phosphorylation were upregulated in animals with H-OA content. Therefore, the enrichment of target genes associated with the AMPK signaling pathway in L-OA might indicate a post-transcriptional regulation of this pathway resulting in downregulated oxidative metabolism in these animals, complementing the findings by Cesar et al. (2016). Still on the turquoise module, the hub bta-miR-146b, which is in the same miRNA family as bta-miR-146a, has been correlated with target genes that are functionally enriched for GO terms associated with fatty acid oxidation (Oliveira et al., 2018).

Target genes of the miRNA blue module were enriched for GO terms associated with proteoglycans in cancer pathways. Proteoglycans have been shown to be key macromolecules that contribute to biology of various types of cancer (Iozzo and Sanderson, 2011). Previous studies have reported an important contribution of OA intake to human health, with protective effects against cancer development (Schwartz et al., 2008). In this sense, genes related to cancer were found to be upregulated in animals with low CLA content (Cesar et al., 2016). These findings may be evidence of miRNA modulation of a carcinogenic pathway. In the miRNA blue module as well, a hub miRNA, bta-miR-21-5p, has been reported to be an important regulator of bovine mammary lipogenesis and metabolism (Li et al., 2015). As in meat, fatty acid composition can influence the nutritional quality of milk and milk fat (Soyeurt et al., 2008). From miRNA yellow module, target genes were enriched for Wnt signaling pathway. Wnt is a member of signal transduction pathways, which regulates crucial aspects of cell fate determination (Komiya and Habas, 2008). It has been reported that the knockdown of a key enzyme in fatty acid synthesis (*FASN*) could attenuate the Wnt signaling pathway *via* downregulation of specific genes (Wang et al., 2016).

Two miRNA modules (red and blue) identified in the H-CLA group were enriched for GO terms associated with insulin signaling pathway and insulin resistance. The hub bta-miR-30b-5p from red module has been reported to regulate muscle cell differentiation (Zhang et al., 2016), while hub bta-miR-10b from blue module and bta-miR-146b from grey module have been reported to have a higher expression in mammary tissue (Wicik et al., 2016). Moreover, bta-miR-146b was associated with the GO terms of Wnt and inflammatory pathways. The canonical Wnt signaling pathway was also enriched in the L-OA group.

In the L-CLA group, target genes in the miRNA blue module were enriched for the insulin signaling pathway and target genes

of the miRNA red module for MAPK signaling pathway GO terms. Besides being involved with inflammatory processes, the MAPK signaling pathway is involved in the activation of PPARα (peroxisome proliferator-activated receptor) by adiponectin, stimulating fatty acid oxidation in muscle cells (Myeong et al., 2006). Adiponectin is an adipocytokine secreted by adipocytes with its beneficial effects on insulin resistance and metabolic disorders (Myeong et al., 2006). Interestingly, hub miRNAs from the red module such as bta-let-7a-5p, bta-let-7f, and bta-let-7e have been associated with insulin-like growth factor receptor signaling pathway (Wicik et al., 2016).

To better understand the biological processes that influence muscle FA composition, PCIT analysis was conducted. In this analysis, we identified the potential regulators that could be involved in the gene expression changes in skeletal muscle due to OA and CLA content. Negative and positive regulators are defined based on the number of significant partial correlations that a gene has between two states (Reverter and Chan, 2008).

Differential hubbing (DH) analysis identified bta-mir-339a as one of the top five negative regulators, with more connections in the L-OA group. The target genes of bta-mir-339a were enriched for 10 significant pathways, including the MAPK signaling pathway, which was also identified by WGCNA. Furthermore, bta-mir 339 has been reported to be expressed at a higher level in bovine adipose tissue than in other tissues (Gu et al., 2007). DH analysis also identified the *FRAT1* gene as a potential negative regulator, which is associated with the Wnt signaling pathway. As discussed previously, the Wnt signaling pathway would be affected by the knockdown of *FASN*, resulting in lower fatty acid synthesis, which complements our findings from WGCNA in the L-OA group. Among the top 10 positive differentially hubbed genes is the *ATP6V0E* gene, from the same family of *ATP6V1D* gene, which has been reported as a factor mediating hepatic steatosis (Nakadera et al., 2016), a metabolic syndrome frequently associated with obesity and diabetes. Furthermore, the *GAL3ST3* gene has been associated with lipid biosynthetic processes (Suzuki et al., 2001).

For CLA content, DH analysis identified *KAT5* as extreme negative and *PSMG1* as extreme positive hubbed genes associated with GO terms for proteasome pathways. The proteasome is a large protein complex responsible for degradation of intracellular proteins (Tanaka, 2009), and proteasome dysfunction has been associated with oxidative stress and insulin sensitivity in human obesity (Díaz-Ruiz et al., 2015). Gonçalves et al. (2018) otherwise concluded that the proteasome pathway may be a potential regulator of beef tenderness in this population. The total lipid content of muscle has a recognized role in beef tenderness, and the concentration of fatty acids is positively correlated with the palatability of beef (Wood et al., 2008). The proteasome pathway was also previously associated with OA content in this population (Cesar et al., 2016), supporting our findings of genes related to proteasome pathways as potential regulators.

Bta-miR-10b was identified as a CLA candidate regulator by both RIF2 and PIF analyses, being also pointed as a hub miRNA by WGCNA. Although its target genes were not enriched for any pathways specifically related to fatty acid metabolism, this miRNA has been previously correlated with backfat thickness and adipose tissue in cattle (Gu et al., 2007; Jin et al., 2010). Bta-mir-486, which was identified by PIF analysis as a top regulator for CLA content, has been associated with skeletal muscle growth (Jing et al., 2015) and was recently found to be downregulated in feed efficient animals of this same cattle population (De Oliveira et al., 2018).

ACTA1 is the gene with the highest PIF rank, which indicates that it may be the most important gene related to both OA and CLA variations in this population. *ACTA1* expression is specific to muscle fibers, with an essential role in muscle contraction and cell morphology (Pollard and Cooper, 1986). *ALDOA* gene was the fourth and second major regulator identified by RIF and PIF analyses for OA and CLA content, respectively. This gene is involved in adipogenic differentiation, which is critical for intramuscular fat deposition and meat quality (Li et al., 2016). The same group of genes (*TPM2*, *CKM*, *TPM1*, and *MYL1*), including *ALDOA*, was identified as positive RIF1 and RIF2 regulators for OA content, indicating the relevance of these genes in fatty acid metabolism. Moreover, Oliveira et al. (2018) also identified the *ALDOA* gene as a putative regulatory for the differences in IMF deposition, which, taken together, may indicate that *ALDOA* is an important gene regulator of fatty acid deposition and composition in Nelore cattle.

In this integrative analysis, insulin resistance, insulin, and MAPK signaling pathways were overrepresented in high and low fatty acid groups. These signaling pathways have been linked to adipocyte differentiation and lipogenesis in cattle (Guo et al., 2017). Based on the literature and our results, insulin and inflammatory processes are influencing OA and CLA composition in Nelore cattle. This study also indicates that hub miRNAs like bta-miR-33a/b, bta-miR-100, bta-miR-204, bta-miR-365-5p, bta-miR-660, bta-miR-411a, bta-miR-136, bta-miR-30-5p, bta-miR-146b, bta-let-7a-5p bta-let-7f, and bta-let-7e are involved with these biological processes. Among the results pointed out by both RIF and PIF analyses, the bta-mir 339, bta-mir-10b, bta-miR 486, and genes *ACTA1* and *ALDOA* are the most relevant regulators for muscle fatty acid composition in Nelore cattle.

Fat and fatty acids, whether in adipose tissue or muscle, contribute to various aspects of meat quality and are central to the nutritional value of meat (Wood et al., 2008). Furthermore, they can have beneficial effects on human health. OA consumption is associated with low levels of low-density lipoprotein (LDL) or "bad cholesterol," which in turn may reduce atherosclerosis risk and diabetes occurrence. Further, OA consumption could increase levels of high-density lipoprotein (HDL) in blood (Smith et al., 2009), whereas CLA consumption may contribute to reduced body fat, cardiovascular diseases, and cancer and can modulate inflammatory responses (Dilzer and Park, 2012).

CONCLUSIONS

In the present study, signaling pathways, miRNAs, and gene regulators related to fatty acid composition in Nelore cattle were identified by miRNA expression and gene co-expression network approaches. Although some of these potential regulators have been previously linked to fatty acid composition, the complex

miRNA–mRNA regulatory network has never been reported so far. This study improves our understanding of the molecular mechanisms controlling intramuscular muscle fat composition in bovines, revealing new candidate networks regulating OA and CLA phenotypes, which could positively benefit beef production.

ETHICS STATEMENT

Experimental procedures were carried out in accordance with the relevant guidelines provided by the Institutional Animal Care and Use Committee Guidelines of the Embrapa Pecuria Sudeste—Protocol CEUA 01/2013. The Ethical Committee of the Embrapa Pecuria Sudeste (Sao Carlos, Sao Paulo, Brazil) approved all experimental protocols (approval code CEUA 01/2013) prior to the conduction of the study.

AUTHOR CONTRIBUTIONS

PO, LC, and LR conceived and designed the experiment; PO, LC, AC, GM, AZ, and LC performed the experiments; PO, AC, WD, and MS performed analysis; PO, AC, WD, BA, JK, JR, and LR interpreted the results; PO, AC, WD, MS, BA, JK, JR, and LR drafted and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was conducted with funding from EMBRAPA, São Paulo Research Foundation (FAPESP, grant number: 2012/23638-8), and scholarship to PO (grant numbers: 2014/22235-1 and 2016/03291-4), the National Council for Scientific and Technological Development (CNPq, grant numbers: 449172/2014-7, 303754/2016-8, 444374/2014-0, and 309004/2016-0), and fellowships to LR, LC, and GM.

ACKNOWLEDGMENTS

We thank the São Paulo Research Foundation (FAPESP) and the National Council for Scientific and Technological Development (CNPq) for providing financial support. We also thank the Iowa State University for accepting the first author as a visiting scholar. We thank Dr. Dante P. D. Lanna, Dr. Michele L. Nascimento, and Dr. Amália S. Chaves for monitoring the feedlots.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00651/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 4, e05005. doi: 10.7554/eLife.05005

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat.* 57 (1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11 (8), R90. doi: 10.1186/gb-2010-11-8-r90

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., et al. (2015). Putative regulatory factors associated with intramuscular fat content. *PLoS One*. 10 (6), e0128350. doi: 10.1371/journal.pone.0128350

Cesar, A. S. M., Regitano, L. C. A., Poleti, M. D., Andrade, S. C. S., Tizioto, P. C., Oliveira, P. S. N., et al. (2016). Differences in the skeletal muscle transcriptome profile associated with extreme values of fatty acids content. *BMC Genomics*. 17 (1), 961. doi: 10.1186/s12864-016-3306-x

Cesar, A. S., Regitano, L. C., Tullio, R. R., Lanna, D. P., Nassu, R. T., Mudado, M., et al. (2014). Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. *BMC Genet.* 15, 39. doi: 10.1186/1471-2156-15-39

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2 (10), 2366–2382. doi: 10.1038/nprot.2007.324

Dávalos, A., Goedeke, L., Smibert, P., Ramírez, C. M., Warrier, N. P., Andreo, U., et al. (2011). miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9232–9237. doi: 10.1073/pnas.1102281108

De Oliveira, P. S. N., Coutinho, L. L., Tizioto, P. C., Cesar, A. S. M., de Oliveira, G. B., Diniz, W. J., et al., et al. (2018). An integrative transcriptome analysis indicates regulatory mRNA–miRNA networks for residual feed intake in Nelore cattle. *Sci. Rep.* 8 (1), 17072. doi: 10.1038/s41598-018-35315-5

Díaz-Ruiz, A., Guzmán-Ruiz, R., Moreno, N. R., García-Rios, A., Delgado-Casado, N., Membrives, A., et al. (2015). Proteasome dysfunction associated to oxidative stress and proteotoxicity in adipocytes compromises insulin sensitivity in human obesity. *Antioxid. Redox Signal.* 23 (7), 597–612. doi: 10.1089/ars.2014.5939

Dilzer, A., and Park, Y. (2012). Implication of conjugated linoleic acid (CLA) in human health. *Crit. Rev. Food Sci. Nutr.* 52 (6), 488-513. doi: 10.1080/10408398.2010.501409

Dimitriadis, G., Mitron, P., Lambadiari, V., Maratou, E., and Raptis, S. A. (2011). Insulin effects in muscle and adipose tissue. *Diabetes Res. Clin. Pract.* Suppl 1, S52–9. doi: 10.1016/S0168-8227(11)70014-6

Fatima, A., and Morris, D. G. (2013). MicroRNAs in domestic livestock. *Physiol. Genomics*. 45 (16), 685–696. doi: 10.1152/physiolgenomics.00009.2013

Filteau, M., Pavey, S. A., St-Cyr, J., and Bernatchez, L. (2013). Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol. Biol. Evol.* 30 (6), 1384–1396. doi: 10.1093/molbev/mst053

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26 (4), 407–415. doi: 10.1038/nbt1394

Gonçalves, T. M., de Almeida Regitano, L. C., Koltes, J. E., Cesar, A. S. M., da Silva Andrade, S. C., Mourão, G. B., et al. (2018). Gene co-expression analysis indicates potential pathways and regulators of beef tenderness in nellore cattle. *Front. Genet.* 9, 441. doi: 10.3389/fgene.2018.00441

Gu, Z., Eleswarapu, S., and Jiang, H. (2007). Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland. *FEBS Lett.* 581 (5), 981–988. doi: 10.1016/j.febslet.2007.01.081

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. 466 (7308), 835–840. doi: 10.1038/nature09267

Guo, Y., Zhang, X., Huang, W., and Miao, X. (2017). Identification and characterization of differentially expressed miRNAs in subcutaneous adipose between Wagyu and Holstein cattle. *Sci. Rep.* 7, 44026. doi: 10.1038/srep44026

Hudson, N. J., Reverter, A., and Dalrymple, B. P. (2009). A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput. Biol.* 5 (5), e1000382. doi: 10.1371/journal.pcbi.1000382

Iozzo, R. V., and Sanderson, R. D. (2011). Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J. Cell. Mol. Med.* 15 (5), 1013–1031. doi: 10.1111/j.1582-4934.2010.01236.x

Jin, W., Dodson, M. V., Moore, S. S., Basarab, J. A., and Guan, L. L. (2010). Characterization of microRNA expression in bovine adipose tissues: a potential

regulatory mechanism of subcutaneous adipose tissue development. *BMC Mol. Biol.* 11, 29. doi: 10.1186/1471-2199-11-29

Jing, L., Hou, Y., Wu, H., Miao, Y., Li, X., Cao, J., et al. (2015). Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential residual feed intake in pigs. *Sci. Rep.* 5, 11953. doi: 10.1038/srep11953

Komiya, Y., and Habas, R. (2008). Wnt signal transduction pathways. *Organogenesis*. 4 (2), 68–75. doi: 10.4161/org.4.2.5851

Laaksonen, D. E., Lindström, J., Lakka, T. A., Eriksson, J. G., Niskanen, L., Wikström, K., et al. (2005a). Physical activity in the prevention of type 2 diabetes: the Finnish Diabetes prevention study. *Diabetes*. 54 (1), 158–165. doi: 10.2337/diabetes.54.1.158

Laaksonen, D. E., Nyyssönen, K., Niskanen, L., Rissanen, T. H., and Salonen, J. T. (2005b). Prediction of cardiovascular mortality in middle-aged men by dietary and serum linoleic and polyunsaturated fatty acids. *Arch. Intern. Med.* 165 (2), 193–199. doi: 10.1001/archinte.165.2.193

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 9, 559. doi: 10.1186/1471-2105-9-559

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25

Lau, D. C. W. (2005). Adipokines: molecular links between obesity and atheroslcerosis. *Am. J. Physiol. Heart Circ. Physiol.* 288 (5), H2031–H2041. doi: 10.1152/ajpheart.01058.2004

Lawless, N., Vegh, P., O'Farrelly, C., and Lynn, D. J. (2014). The role of microRNAs in bovine infection and immunity. *Front. Immunol.* 5, 611. doi: 10.3389/fimmu.2014.00611

Li, R., Beaudoin, F., Ammah, A. A., Bissonnette, N., Benchaar, C., Zhao, X., et al. (2015). Deep sequencing shows microRNA involvement in bovine mammary gland adaptation to diets supplemented with linseed oil or safflower oil. *BMC Genomics*. 16, 884. doi: 10.1186/s12864-015-1965-7

Li, X., Huang, K., Chen, F., Li, W., Sun, S., Shi, X. E., et al. (2016). Verification of suitable and reliable reference genes for quantitative real-time PCR during adipogenic differentiation in porcine intramuscular stromal-vascular cells. *Animal*. 10 (6), 947–952. doi: 10.1017/S1751731115002748

Lian, S., Guo, J. R., Nan, X. M., Ma, L., Loor, J. J., and Bu, D. P. (2016). MicroRNA Bta-miR-181a regulates the biosynthesis of bovine milk fat by targeting ACSL1. *J. Dairy Sci.* 99 (5), 3916–3924 doi: 10.3168/jds.2015-10484

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Moloney, F., Toomey, S., Noone, E., Nugent, A., Allan, B., Loscher, C. E., et al. (2007). Antidiabetic effects of cis-9, trans-11-conjugated linoleic acid may be mediated via anti-inflammatory effects in white adipose tissue. *Diabetes*. 56 (3), 574–582. doi: 10.2337/db06-0384

Myeong, J. Y., Gha, Y. L., Chung, J. J., Young, H. A., Seung, H. H., and Jae, B. K. (2006). Adiponectin increases fatty acid oxidation in skeletal muscle cells by sequential activation of AMP-activated protein kinase, p38 mitogen-activated protein kinase, and peroxisome proliferator-activated receptor{alpha}. *Diabetes*. 55 (9), 2562–2570. doi: 10.2337/db05-1322

Nakadera, E., Yamashina, S., Izumi, K., Inami, Y., Sato, T., Fukushima, H., et al. (2016). Inhibition of mTOR improves the impairment of acidification in autophagic vesicles caused by hepatic steatosis. *Biochem. Biophys. Res. Commun.* 469 (4), 1104–1110. doi: 10.1016/j.bbrc.2015.12.010

Nakamura, M. T., and Nara, T. Y. (2003). Essential fatty acid synthesis and its regulation in mammals. *Prostaglandins Leukot. Essent. Fatty Acids*. 68 (2), 145–150. doi: 10.1016/S0952-3278(02)00264-8

Oliveira, G. B., Regitano, L. C. A., Cesar, A. S. M., Reecy, J. M., Degaki, K. Y., Poleti, M. D., et al. (2018). Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics*. 19 (1), 126. doi: 10.1186/s12864-018-4514-3

Ortega, F. J., Mercader, J. M., Catalán, V., Moreno-Navarrete, J. M., Pueyo, N., Sabater, M., et al. (2013). Targeting the circulating microRNA signature of obesity. *Clin. Chem.* 59 (5), 781–792. doi: 10.1373/clinchem.2012.195776

Pollard, T. D., and Cooper, J. A. (1986). Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. *Annu. Rev. Biochem.* 55, 987–1035. doi: 10.1146/annurev.bi.55.070186.005011

Reverter, A., and Chan, E. K. F. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*. 24 (21), 2491–2497. doi: 10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Pérez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics*. 26 (7), 896–904. doi: 10.1093/bioinformatics/btq051

Romao, J. M., Jin, W., He, M., McAllister, T., and Guan, L. L. (2014). MicroRNAs in bovine adipogenesis: genomic context, expression and function. *BMC Genomics*. 15, 137. doi: 10.1186/1471-2164-15-137

Scaglia, N., and Igal, R. A. (2005). Stearoyl-CoA desaturase is involved in the control of proliferation, anchorage-independent growth, and survival in human transformed cells. *J. Biol. Chem.* 280 (27), 25339–25349. doi: 10.1074/jbc.M501159200

Schwartz, G. J., Fu, J., Astarita, G., Li, X., Gaetani, S., Campolongo, P., et al. (2008). The lipid messenger OEA links dietary fat intake to satiety. *Cell Metab.* 8 (4), 281–288. doi: 10.1016/j.cmet.2008.08.005

Smith, S. B., Lunt, D. K., Chung, K. Y., Choi, C. B., Tume, R. K., and Zembayashi, M. (2006). Adiposity, fatty acid composition, and delta-9 desaturase activity during growth in beef cattle. *Anim. Sci. J.* 77 (5), 478–486. doi: 10.1111/j.1740-0929.2006.00375.x

Smith, S., Gill, C., Lunt, D., and Brooks, M. (2009). Regulation of fat and fatty acid composition in beef cattle. *Asian-australas. J. Anim. Sci.* 22 (9), 1225–1233. doi: 10.5713/ajas.2009.r.10

Soares-Silva, M., Diniz, F. F., Gomes, G. N., and Bahia, D. (2016). The mitogen-activated protein kinase (MAPK) pathway: role in immune evasion by trypanosomatids. *Front. Microbiol.* 7, 183. doi: 10.3389/fmicb.2016.00183

Soyeurt, H., Dardenne, P., Dehareng, F., Bastin, C., and Gengler, N. (2008). Genetic parameters of saturated and monounsaturated fatty acid content and the ratio of saturated to unsaturated fatty acids in bovine milk. *J. Dairy Sci.* 91 (9), 3611–3626. doi: 10.3168/jds.2007-0971

Suzuki, A., Hiraoka, N., Suzuki, M., Angata, K., Misra, A. K., McAuliffe, J., et al. (2001). Molecular cloning and expression of a novel human β-Gal-3-O-sulfotransferase that acts preferentially on N-acetyllactosamine in N- and O-glycans. *J. Biol. Chem.* 276 (26), 24388–24395. doi: 10.1074/jbc.M103135200

Tanaka, K. (2009). The proteasome: overview of structure and functions. *Proc. Jpn. Acad., Ser. B*. 85 (1), 12–36. doi: 10.2183/pjab.85.12

Thomson, D. M., Herway, S. T., Fillmore, N., Kim, H., Brown, J. D., Barrow, J. R., et al. (2008). AMP-activated protein kinase phosphorylates transcription factors of the CREB family. *J. Appl. Physiol.* 104 (2), 429–438. doi: 10.1152/japplphysiol.00900.2007

Valsta, L. M., Tapanainen, H., and Männistö, S. (2005). Meat fats in nutrition. *Meat Sci.* 70 (3), 525–530. doi: 10.1016/j.meatsci.2004.12.016

Wang, H., Xi, Q., and Wu, G. (2016). Fatty acid synthase regulates invasion and metastasis of colorectal cancer via Wnt signaling pathway. *Cancer Med.* 5 (7), 1599–1606. doi: 10.1002/cam4.711

Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 45 (W1), W130–W137. doi: 10.1093/nar/gkx356

Wicik, Z., Gajewska, M., Majewska, A., Walkiewicz, D., Osińska, E., and Motyl, T. (2016). Characterization of microRNA profile in mammary tissue of dairy and beef breed heifers. *J. Anim. Breed. Genet.* 133 (1), 31–42. doi: 0.1111/jbg.12172

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., et al. (2008). Fat deposition, fatty acid composition and meat quality: a review. *Meat Sci.* 78 (4), 343–358 doi: 10.1016/j.meatsci.2007.07.019

Zhang, W. W., Sun, X. F., Tong, H. L., Wang, Y. H., Li, S. F., Yan, Y. Q., et al. (2016). Effect of differentiation on microRNA expression in bovine skeletal muscle satellite cells by deep sequencing. *Cell. Mol. Biol. Lett.* 21, 8. doi: 10.1186/s11658-016-0009-x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 de Oliveira, Coutinho, Cesar, Diniz, de Souza, Andrade, Koltes, Mourão, Zerlotini, Reecy and Regitano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 July 2019
doi: 10.3389/fgene.2019.00661

Differential microRNA Expression in Porcine Endometrium Involved in Remodeling and Angiogenesis That Contributes to Embryonic Implantation

*Linjun Hong[1,2], Ruize Liu[3], Xiwu Qiao[1,2], Xingwang Wang[1,2], Shouqi Wang[1,2], Jiaqi Li[1,2], Zhenfang Wu[1,2] and Hao Zhang[1,2]**

[1] National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou, China, [2] Guangdong Provincial Key Laboratory of Agro-animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou, China, [3] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, United States

OPEN ACCESS

Edited by:
David E MacHugh,
University College Dublin, Ireland

Reviewed by:
Tao Zhou,
Auburn University, United States
Xiangdong Ding,
China Agricultural University (CAU), China

***Correspondence:**
Hao Zhang
zhanghao@scau.edu.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 17 February 2019
Accepted: 24 June 2019
Published: 26 July 2019

Citation:
Hong L, Liu R, Qiao X, Wang X, Wang S, Li J, Wu Z and Zhang H (2019) Differential microRNA Expression in Porcine Endometrium Involved in Remodeling and Angiogenesis That Contributes to Embryonic Implantation. Front. Genet. 10:661. doi: 10.3389/fgene.2019.00661

Background: In western swine breeds, up to 30% of embryonic losses occur during early pregnancy, and the majority of embryonic losses happens during implantation. In this period, maternal recognition of pregnancy begins to occur and blastocysts undergo dramatic morphologic changes. As with other species, changes in the uterine environment plays an important role in the process of embryo implantation in pigs. Erhualian (ER) pigs, one of the Chinese Taihu swine breeds, are known to have the highest litter size in the world. Experiments demonstrated that the greater embryonic survival on gestation day (GD) 12 in Chinese Taihu pigs is one important factor that contributes to enhanced litter size. This is largely controlled by maternal genes. In this study, endometrial samples were collected from pregnant Landrace×Large Yorkshire (LL) sows (parity 3) and ER sows (parity 3) on GD12 and the expression profiles of microRNAs (miRNAs) in the endometrium were compared between ER and LL using miRNA-seq technology.

Results: A total of 288 miRNAs were identified in the pig endometrium, including 202 previously known and 86 novel miRNAs. The Kyoto Encyclopedia of Genes and Genomes pathway analysis revealed that highly abundant miRNAs might affect endometrial remodeling. Comparison between LL and ER sows revealed that 96 known miRNAs were differentially expressed between the two groups (including 78 up-regulated and 18 down-regulated miRNAs in ER compared to LL). Bioinformatics analysis showed that the target genes of some differentially expressed miRNAs were involved in pathways related to angiogenesis, proliferation, apoptosis, and tissue remodeling, which play critical roles in implantation by regulating endometrial structural changes and secretions of hormones, growth factors, and nutrients. Furthermore, the results demonstrated that insulin-like growth factor-1 protein expression was directly inhibited by miR-206. The lower expression of miR-206 in ER compared to LL might facilitate the angiogenesis of the endometrium during embryo implantation.

Conclusions: The identified miRNAs that are differentially expressed in the endometrium of ER and LL pigs will contribute to the understanding of the role of miRNAs in embryonic implantation and the molecular mechanisms of the highest embryonic survival in Chinese ER pigs.

Keywords: porcine, endometrium, microRNAs, differential expression, implantation

INTRODUCTION

Litter size has a great impact on the profitability of swine production. Prenatal mortality is the major limitation for increasing the litter size in pigs. Up to 30% of conceptuses are spontaneously lost during early pregnancy, especially on gestation days (GD) 11 to 13 (Scofield et al., 1974; Pope and First, 1985; Zavy and Geisert, 1994; Wilson et al., 1999). In contrast to rodents and primates, pigs have an extended period of preimplantation. From GD4 to GD12, the developing conceptuses undergo speedy morphologic changes (from spherical to tubular to filamentous forms) and migrate freely in the uterine cavity. On GD15, filamentous conceptuses grow to 800 to 1,000 mm in length and begin to attach to luminal uterine epithelium (LE) (Geisert et al., 1982; Bazer and Johnson, 2014). Thus, during protracted preimplantation, the requirement for nutrients of conceptuses is mainly dependent on uterine secretions, including glucose, amino acids, ions, enzymes, growth factors, hormones, growth factors, and other substances termed as histotroph (Spencer et al., 2006).

Chinese Taihu pigs, including Erhualian (ER), Meishan, and Fengjing breeds, are highly prolific. ER pigs are known to have the biggest litter size record in the world (Zhang, 1986). Meishan pigs were exported to western countries in the early 1980s and have been studied for more than 30 years to explore the mechanism of prolificacy. Studies found that the greater embryonic survival on GD11 to GD12 in Chinese Taihu pig is the most important factor contributing to enhanced litter size, and this is controlled primarily by maternal genes (Haley et al., 1995). At this stage, substantial changes occur in the conceptus-uterine interface, including morphologic changes in the conceptus, the onset of synthesis of estradiol by the conceptus, and the appropriate physiologic adjustments of uterus (Bazer and Johnson, 2014). Examination of individual embryos in Meishan pigs found that embryo survival was 108.1% on GD11 and 93.3% on GD12; however, in Landrace×Large Yorkshire (LL) pigs, embryo survival was 89.1% on GD11 and 49.9% on GD12 (Ashworth et al., 1997). Other studies also demonstrated that, from GD11 to GD12, the embryonic survival rate in Chinese Taihu pigs was significantly higher than in western pigs (Bazer et al., 1988; Christenson et al., 1993). Thus, it is worth further studying the molecular mechanisms underlying the differences in uterine environment changes between Chinese Taihu and western pigs.

microRNAs (miRNAs) are short (20–25 nt), endogenous, conserved, non-coding RNA molecules that play wide biological roles in transcription and translation (Carrington and Ambros, 2003; Bartel, 2004). miRNAs typically interact with target mRNAs by base pairing and destabilize or degrade their complementary mRNA (Wu et al., 2006). They have been shown to participate in the regulation of various physiologic processes, including cellular proliferation, differentiation, apoptosis, angiogenesis, embryonic development, and reproduction control (Laurent, 2010; Nicoli et al., 2012; Rosenbluth et al., 2013). A large number of miRNAs have been shown to be associated with embryo implantation in humans and mice, such as the regulation of endometrial receptivity (Altmäe et al., 2013) and endometrial stromal cell differentiation (Qian et al., 2009), participating in human pregnancy and parturition (Montenegro et al., 2009). In pigs, Su et al. and Liu et al. reported that miRNAs play roles in porcine placental growth and functions (Su et al., 2010; Liu et al., 2015). In addition, Wessels et al. investigated the expression of miRNAs on both sides of the maternal-fetal interface in the model of implantation failure and spontaneous fetal loss in pigs and identified miRNAs that might contribute to fetal loss (Wessels et al., 2013). Thus, taken together, these results indicate the importance of miRNAs in pig reproduction. In this study, the expression profiles of miRNAs in the sow endometrium on GD12 were compared between ER and LL pigs using sequencing technology. Differentially expressed miRNAs (DEMs) were identified and miRNAs involved in reproduction were analyzed by bioinformatic analysis and experiments. Collectively, these results will help better understand the role of miRNAs in embryonic survival during implantation.

MATERIALS AND METHODS

Tissue Collection

All of the experiments involving animals were conducted according to animal ethics guidelines and approved by the Animal Care and Use Committee of South China Agricultural University (Guangzhou, China). LL and ER sows were obtained from the breeding pig farm of Guangdong Wen's Foodstuffs Group Co., Ltd. (Yunfu, China). Three LL sows (parity 3) and three ER sows (parity 3) were checked for estrus twice daily and artificially inseminated at the onset of estrus (day 0) and again 12 h later. After the sows were slaughtered at a local slaughterhouse on GD12, the uteri were removed rapidly and transported in an icebox to the laboratory. Pregnancy was confirmed by the presence of apparently normal filamentous conceptuses in uterine flushings. Endometrial samples were collected and stored at −80°C for RNA extraction.

RNA Extraction and Small RNA (sRNA) Sequencing

Total RNA was extracted from six endometrial samples using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instruction. RNA purity was quantified using

NanoDrop ND2000 spectrophotometer at 260 and 280 nm (Thermo Fisher Scientific, Wilmington, MA, USA), and RNA integrity was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The OD260/OD280 ratios of all the samples were greater than 1.8, and the RIN values were greater than 8. Equal RNA quantities from the endometria of three pigs from the LL and ER groups were pooled. sRNA Illumina sequencing was conducted as follows: ~10 μg total RNA was size fractionated by Novex 15% TBE-Urea gel and RNA fragments between 18 and 30 bases in length were isolated. The purified sRNAs were then ligated with the 5′-adapter. To remove unligated adapters, the ligation products (36–50 bases in length) were gel purified on Novex 15% TBE-Urea gel. Subsequently, the RNA fragments with the adapter at the 5′-end were ligated with 3′-adapters. After gel purification on Novex 10% TBE-Urea gel, RNA fragments with adapters at both ends (62–75 bases long) were reverse transcribed. Reverse transcription-polymerase chain reaction (RT-PCR) was used to create cDNA constructs based on the sRNA ligated with the 5′- and 3′-adapters. This protocol gel purifies the amplified cDNA construct in preparation for loading on the Illumina Cluster Station. The cDNAs were amplified using the appropriate PCR cycles to produce sequencing libraries. Sequencing was carried out at BGI-Shenzhen, China.

Sequence Analysis

First, raw data (raw reads) were processed by custom Perl and Python scripts, raw reads contain poly-A/T/G/C, poly-N, with 5′-adapter contaminants, without 3′-adapter, or the insert tag, and low-quality reads were filtered to get clean data. Subsequently, clean reads ≥18 nt were chosen as sRNA tags and mapped to reference sequence by Bowtie (Langmead et al., 2009) without mismatch to analyze their expression and distribution on the reference. Third, miRBase (release 20.0) was used as reference, and srna-tools-cli and modified software mirdeep2 (Friedländer et al., 2011) were used to obtain the potential miRNA and draw the secondary structures. The available software mirdeep2 (Friedländer et al., 2011) and miREvo (Wen et al., 2012) were integrated to predict novel miRNAs by exploring the secondary structure. At the same time, custom scripts were used to obtain the identified miRNA counts as well as base bias on the first position with certain length and on each position of all identified miRNAs, respectively.

Identification of DEMs

The procedures that determine the DEMs between LL and ER groups are shown below:

miRNA expression levels were estimated by transcript per million (TPM) with the following criteria (Wagner et al., 2012): Normalization formula: Normalized expression = Actual miRNA count/Total count of clean reads*1,000,000. miRNAs with a normalized expression level of less than 1 in each of the two libraries and miRNAs with an estimated probability value of less than 0.95 were removed. The fold change in the expression level and the *P* value between two libraries were calculated from the normalized expression using the following formulas, respectively:

Fold change formula: Fold change = $\log_2$(ER/LL)

P-value formula:

$$p(x \mid y) = \left(\frac{N_2}{N_1}\right)^y \frac{(x+y)!}{x!\,y!\left(1+\frac{N_2}{N_1}\right)^{(x+y+1)}}$$

N_1 and x represent the total count of clean reads and the normalized expression level of a given miRNA in an sRNA library of LL endometrial tissue sample, respectively. N_2 and y represent the total count of clean reads and the normalized expression level of a given miRNA in an sRNA library of ER endometrial tissue sample, respectively (Audic and Claverie, 1997). Raw *P* values were converted to adjusted *P* values using the Benjamini–Hochberg false discovery rate (Benjamini and Hochberg, 1995). The adjusted $P < 0.05$ and $|\log_2(\text{fold change})| > 1$ were set as thresholds for significantly differential expression by default.

Target Gene Prediction and Functional Analysis of DEMs

The prediction of the target genes of miRNAs was performed by RNAhybrid and TargetScan. Overlapping target genes were selected for further analysis. To reveal the target genes' potential biological functions and identify the main pathways targeted by the gene candidates, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed as described previously (Zhang et al., 2013). Based on the GO and KEGG database, the hypergeometric test was preformed to identify significantly enriched GO terms ($Q < 0.05$) and classify the pathway category (Boyle et al., 2004). The network of pathways based on the GO and KEGG database was constituted by ClueGO, which is a plugin in Cytoscape (http://www.cytoscape.org/).

Validation of miRNA Expression via Stem-Loop Quantitative RT-PCR (qRT-PCR)

The sRNA-seq results were validated using RNA samples from the LL ($n = 3$) and ER ($n = 3$) groups by the stem-loop qRT-PCR method. A total of 16 miRNAs were selected for qRT-PCR validation. The mature miRNA and primer sequences are available in **Supplementary Table S1**. Briefly, for RT-PCR, the Revert Aid™ First Strand cDNA Synthesis Kit (Promega, Fitchburg, WI, USA) was adopted according to the manufacturer's instructions. Then, RT-PCR was performed with SYBR® Premix Ex Taq™ (Toyobo) on ABI PRISM® 7500 Sequence Detection System. Porcine U6 snRNA was used as an internal control and all reactions were run in triplicate. PCR profiles were one cycle at 95°C for 5 min followed by 40× (95°C for 15 s, 65°C for 15 s, and 72°C for 32 s). The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. Fold change ($\log_2$ ratio) was used to show the differential expression of miRNA in LL and ER.

Dual-Luciferase Reporter Assays

For luciferase reporter experiments, the pmirGLO dual-luciferase reporter vector (Promega) housing the 3′-untranslated

region (UTR) of insulin-like growth factor-1 (IGF-1), which was *Xho*I and *Xba*I cloned to the 3′-end of the *Renilla* gene, was used to examine the effect of miR-206 on *Renilla* production. IGF-1 3′-UTR vector (pmirGLO-IGF1) containing the miR-206 binding site (CATTCC) was constructed by RT-PCR using specific primers (forward primer 5′-CCGCTCGAGCAGGAAACAAGA ACTACAG-3′ and reverse primer 5′-GCTCTAGACAACAGCA ATCTACCAACT-3′). Meanwhile, IGF-1 3′-UTR-Mutant vector (pmirGLO-IGF1-Mutant) with a mutated miR-206 binding site (GTAAGG) was also constructed. The miR-206 mimics and its mutant mimics (miR-206_mut) were designed and synthesized by GenePharma Biotech Co. (Shanghai, China). In the dual-luciferase assays, PK15 cells were cultured in DMEM complete medium (Hyclone, Logan, UT, USA) and then plated onto a 96-well plate. The miR-206 mimics, mutant miR-206mimics, or negative control (NC) were co-transfected into cells with 3′-UTR dual-luciferase vector using Lipofectamine 2000 (Invitrogen, Shanghai, China). Cells were collected 24 h after transfection, and assayed with the Dual-Luciferase Reporter Assay System (Promega). Three replicates were performed for each transfection.

Lentivirus Preparation and Administration

The pri-miR-206 expression lentivirus vector (H1-MCS-CMV-EGFP) and the NC lentivirus vector were purchased from GenePharma Biotech. Virus titration and infection efficiency were measured by the fluorescence method as lentiviral vectors expressed enhanced green fluorescent protein in infected cells. According to the results of a preliminary experiment, the titer of lenti-pri-miR-206 used for experiments was 1×10^7 TU/ml. The lentivirus vectors were transfected into porcine skeletal muscle satellite cells (SCs) with a titer of 1×10^7 TU/ml in the presence of polybrene (5 μg/ml). Cells were collected 72 h after transfection, and total RNA and protein were extracted for further experiments.

Western Blot Analysis

Protein lysates were generated using the mammalian protein extraction reagent RIPA (Beyotime, Shanghai, China). The concentration of extracted total protein from each sample was calculated using the BCA Protein Assay Kit (Thermo Pierce, Rockford, IL, USA). The equivalent protein for each sample was loaded into a 10% sodium docedyl sulfate-polyacrylamide gel electrophoresis and fractionated, and the denatured proteins were subsequently transferred from gel to a polyvinylidene fluoride membrane (Millipore, Billerica, MA, USA) by a Mini-PROTEAN Tube Cell instrument (Bio-Rad, Hercules, CA, USA). The membranes were incubated with antibodies (IGF-1, ab9572, Abcam; glyceraldehyde 3-phosphate dehydrogenase, ab8245, Abcam) overnight at 4°C and then with horseradish peroxidase-conjugated goat anti-rabbit secondary antibody for 1 h at room temperature. The enhanced chemiluminescence substrate (Beyotime) was used to visualize the band, and a picture was captured by an imaging system (UVP, Upland, CA, USA). Finally, the quantification analysis was performed by ImageJ 1.45 software (NIH Image).

Statistical Analysis

Data from the results of qRT-PCR, dual-luciferase reporter assays, and Western blot analysis were analyzed using SPSS version 18.0 (SPSS, Inc., Chicago, IL, USA). Paired *t* tests and two-way analyses of variance were performed to analyze the relative expression of miRNAs, the luciferase activity, and the intensity of the protein band in Western blot analysis. $P < 0.05$ was considered statistically significant.

RESULTS

Overview of the Squences Generated by Illumina Sequencing

sRNA libraries were generated from a total of six samples from ER and LL sows on GD12. After removing low-quality reads and adaptor sequences, a total of 9,104,438 and 12,881,211 clean reads were obtained from ER and LL samples, respectively. The sRNA annotation is presented in **Supplementary Table S2**. The results of sRNA annotation showed that known miRNAs accounted for 48.64% and 56.96% of the total clean reads in ER and LL, respectively (**Supplementary Table S2**). The distribution of sequence lengths was similar between ER and LL libraries (**Figure 1A**). The number of 20 to 23 nt sequences was significantly greater than that of shorter or longer sequences, and almost half of the sequences in LL (47.03%) and 39.33% sequences in ER are 22 nt.

In total sRNA reads, 20,689,836 common sequences were obtained in LL and ER, accounting for 94.11% of the total sequence reads in the two libraries (**Figure 1B**). In unique sRNA reads, 63,548 common sequences were obtained in LL and ER, accounting for 7.69% of the total reads in the two libraries (**Figure 1C**), and 361,418 (43.71%) and 401,914 (48.61%) specific sequences were obtained from LL and ER, respectively (**Figure 1C**).

Sequence Variants and Editing of Bases in the Seed Region of miRNAs

Sequencing data analysis revealed that the majority of identified miRNAs showed length and sequence heterogeneity in the porcine endometrial tissue. The length variations occurred largely in the 3′-end of miRNAs, mainly in the form of terminal reductions or additions of nucleotides. In ER, miR-128, miR-187, miR-18b, miR-190, miR-196a, miR-206, miR-215, miR-2476, miR-326, miR-338, miR-676, and miR-758 had variants only at the 3′-end, whereas, in LL, miR-105, miR-129b, miR-149, miR-153, miR-190, miR-208b, miR-216, miR-450a, miR-450c-5p, miR-499-5p, miR-503, and miR-95 had variants only at the 3′-end. In addition, 11 and 10 miRNAs in the ER and LL libraries, respectively, were mutated by only one nucleotide in the 5′-end, but they had several 3′-end variants (**Supplementary Tables S3** and **S4**). Similarly, previous studies also revealed the length variations of miRNAs in other porcine tissues (Li et al., 2010; Nielsen et al., 2010; Li et al., 2011). Such variants might be from altered miRNA processing, prioritized degradation at miRNA ends, or post-transcriptional modifications, including RNA editing (Aravin and Tuschl, 2005). These end-sequence variations are interesting as they may allow

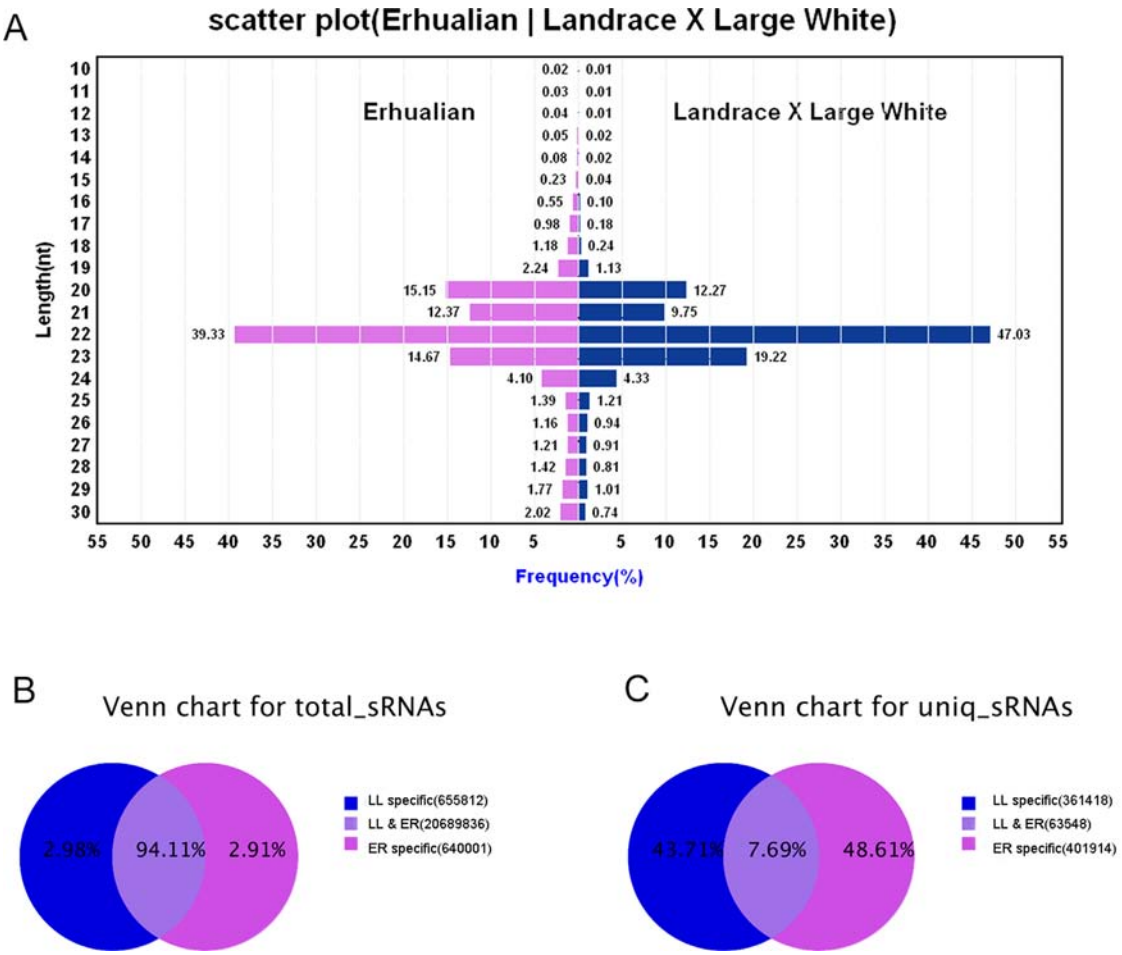


FIGURE 1 | Overview of the sequences generated by Illumina sequencing. **(A)** Sequence length distributions of the two libraries. The length distributions peaked at 22 nt, which is expected for miRNAs' length. **(B)** Number of total sRNA tags between the two libraries. **(C)** Number of unique sRNA tags between the two libraries.

miRNA variants to play different roles by influencing the miRNA-target mRNA hybrid duplex structure (Jazdzewski et al., 2009). The nucleotides at positions 2 to 8 of a mature miRNA are known as the seed region. The seed region binds to a target site in the 3′-UTR of the target mRNA by complementarities and is highly conserved. The target of an miRNA may alter due to change in the nucleotides in the seed region. Editing of bases in the seed region of miRNAs has been reported to occur frequently (Kawahara et al., 2007; Liu et al., 2008). In the present analysis, miRNAs that might have seed editing can be distinguished by matching unannotated sRNA with porcine mature miRNAs from miRBase 20.0. Forty-nine and 62 mature miRNAs in ER and LL had a single nucleotide substitution in the seed region, respectively (**Supplementary Table S5**). The observed occurrence for each possible substitution is summarized in **Supplementary Table S5** for ER and LL samples, respectively. In ER, the most frequent substitutions were T-to-C (22.2%), A-to-G (20.7%), and G-to-A (13.6%), whereas, in LL, the most frequent substitutions were T-to-G (18.1%), T-to-A (15.0%), and G-to-T (14.5%; **Figure 2**). Although the most frequent substitutions were different between ER and LL libraries, C-to-A (0.4% in ER; 2.%) and C-to-G (0.4% in ER; 2.%) were the substitutions with the lowest frequency in both ER and LL libraries. In porcine adipose tissue samples, similar results were also reported (Li et al., 2011). Interestingly, abundant miRNAs (ssc-let-7a, ssc-mir-143, ssc-let-7f, ssc-mir-21, and ssc-mir-378) also had higher editing probability (**Supplementary Table S5**). This indicates that highly expressed miRNAs targeted more genes.

Expression Profiling of miRNAs

In the second part of the present analysis, the global expression profile of endometrial miRNAs on GD12 in LL and ER pigs was determined. A total of 288 miRNAs were identified in the pig endometrium, including 202 known miRNAs and 86 novel miRNAs (**Supplementary Table S6**). Among the known miRNAs, 200 miRNAs were co-expressed in LL and ER and 2 miRNAs (ssc-miR-124a and ssc-miR-450c-3p) were specifically expressed in ER. Among the novel miRNAs, 38 miRNAs were co-expressed in LL and ER and 19 and 28 miRNAs were specifically expressed in LL and ER, respectively.

The 20 most highly expressed miRNAs in LL and ER libraries are listed in **Table 1**. Among them, 14 highly expressed miRNAs

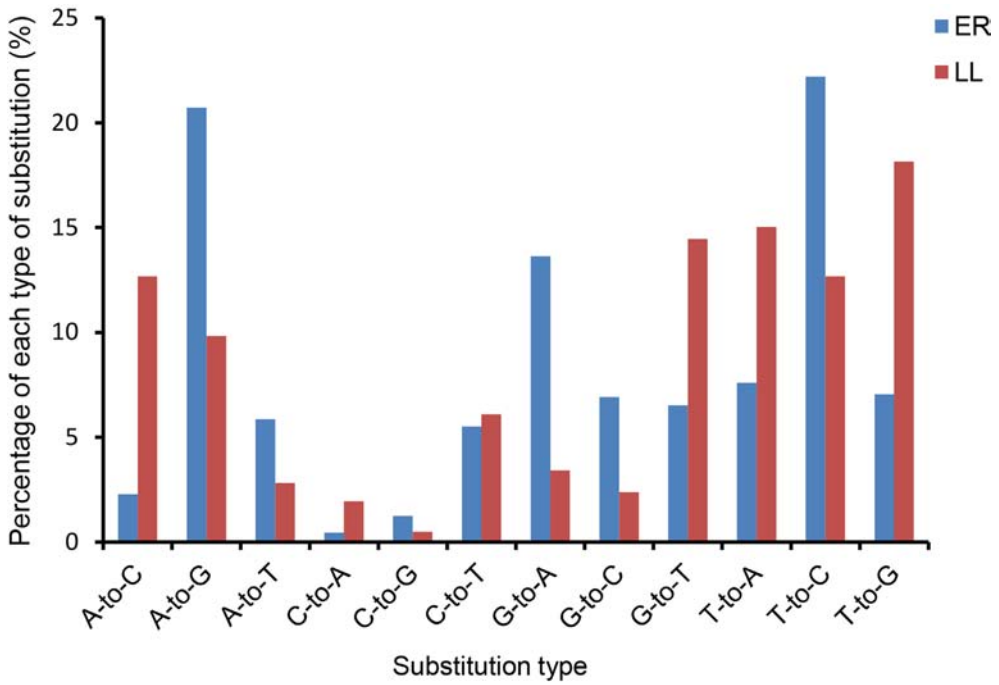


FIGURE 2 | Histogram displaying the single nucleotide substitution in the miRNA seed region sequence when aligning un-annotated sRNA tags with porcine mature miRNAs from miRBase 20.0.

TABLE 1 | Top 20 miRNAs in LL and ER.

miRNA-name	Average TPM in LL	Rank in LL	miRNA-name	Average TPM in ER	Rank in ER
ssc-miR-143-3p	2208321	1	ssc-miR-143-3p	1266790	1
ssc-let-7a	1289290	2	ssc-let-7a	836090	2
ssc-miR-21	519081	3	ssc-miR-21	509408	3
ssc-let-7f	467190	4	ssc-let-7f	317357	4
ssc-miR-30a-5p	313695	5	ssc-miR-30a-5p	92442	9
ssc-miR-148a	290344	6	ssc-miR-148a	31042	18
ssc-miR-10a	242124	7	ssc-miR-10a	112910	8
ssc-miR-10b	240246	8	ssc-miR-10b	133656	6
ssc-let-7c	224940	9	ssc-let-7c	129079	7
ssc-miR-378	184397	10	ssc-miR-378	213166	5
ssc-miR-30d	138588	11	ssc-miR-30d	63006	10
ssc-miR-34c	102496	12	ssc-miR-34c	13563	31
ssc-miR-140*	88368	13	ssc-miR-140*	21112	22
ssc-miR-103	79372	14	ssc-miR-103	32711	17
ssc-let-7g	78468	15	ssc-let-7g	38160	16
ssc-miR-126	45626	16	ssc-miR-126	23587	21
ssc-miR-1	44500	17	ssc-miR-1	4031	57
ssc-miR-191	44154	18	ssc-miR-191	9654	36
ssc-miR-30e-5p	43404	19	ssc-miR-30e-5p	27545	20
ssc-miR-206	41179	20	ssc-miR-206	233	128
ssc-miR-26a	40215	22	ssc-miR-26a	47934	11
ssc-miR-101	34754	23	ssc-miR-101	41815	12
ssc-miR-196b-5p	30481	27	ssc-miR-196b-5p	27714	19
ssc-miR-199a*	21283	32	ssc-miR-199a*	39465	13
ssc-miR-199a-3p	21236	33	ssc-miR-199a-3p	39427	14
ssc-miR-199b*	21236	34	ssc-miR-199b*	39426	15

The list shows the top 20 abundant miRNAs in LL and ER, respectively.
The star is part of the miRNA name.

were the same in LL and ER. Thus, the predicted target genes of 14 common miRNAs were chosen for functional analysis. GO analysis (ClueGo network of GO terms) showed that they were mainly involved in the "cellular protein metabolic process," "regulation of macromolecule biosynthetic process," and "anatomical structure morphogenesis" (**Supplementary Figure S1**). The KEGG pathway analysis (ClueGo network of pathways) indicated that the predicted target genes were mainly enriched in "Apoptosis," "Autophagy," "Ubiquitin-mediated proteolysis," "Longevity-regulating pathway," "AMPK signaling pathway," "Regulation of actin cytoskeleton," "Focal adhesion," "ECM-receptor interaction," "Rap1 signaling pathway," "FoxO signaling pathway," "mTOR signaling pathway," and "MAPK signaling pathway" (**Figure 3**).

Comparative Analysis of DEMs Between ER and LL

These DEMs between LL and ER libraries are listed in **Supplementary Table S7**, and in total, 96 known and 68 novel significantly DEMs were identified between LL and ER groups. Of the 96 differentially expressed known miRNAs, 78 were up-regulated and 18 were down-regulated in ER compared to LL (**Figure 4A**). miR-206 ranked the top [fold change $\log_2$(ER/LL) = −6.96] among DEMs that were expressed in both LL and ER. Of the differentially expressed novel miRNAs, 43 were up-regulated and 25 were down-regulated in ER than in LL (**Figure 4B**).

Validation of Sequencing Results by qRT-PCR

The stem-loop qRT-PCR assay was used to specifically detect mature miRNAs. U6 snRNA was selected as the reference gene. Sixteen miRNAs were chosen for validation by qRT-PCR and the primers used are listed in **Supplementary Table S1**. The expression patterns for the 16 miRNAs were consistent with those in sequencing data (**Figure 5**).

Functional Annotation of DEMs in Endometrial Tissue Samples

To evaluate the biological functions of these DEMs, target genes of DEMs were predicted by RNAhybrid and TargetScan, and the KEGG pathway analysis of these target genes was performed. Thus, 275 significantly enriched signaling pathways were obtained, such as with "VEGF signaling pathway" (Angiogenesis-related pathway), "Toll-like receptor signaling pathway" (Immune-related pathway), "Regulation of actin cytoskeleton" (Tissue remodeling-related pathway), and "MAPK signaling pathway, TGF-β signaling pathway and Apoptosis" (Proliferation and apoptosis) in the top 25 signaling pathways (**Figure 6**). In addition, some other target genes were annotated to reproduction-associated pathways, including "Steroid hormone biosynthesis," "Progesterone-mediated oocyte maturation," "Steroid biosynthesis," "GnRH signaling pathway," and "p53 signaling pathway."

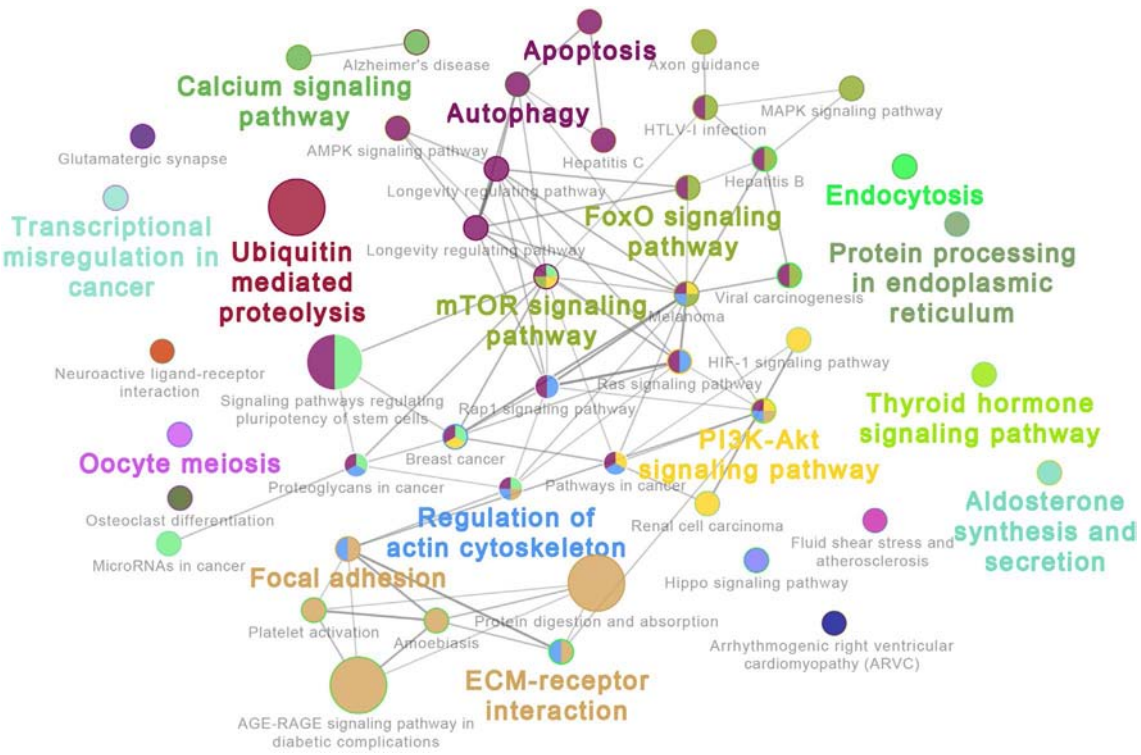


FIGURE 3 | ClueGo network of pathways. Each node represents a pathway. The enrichment significance of pathway is reflected by the size of the nodes. Node color represents the class that they belong to. Mixed coloring means that the specific node belongs to multiple classes.

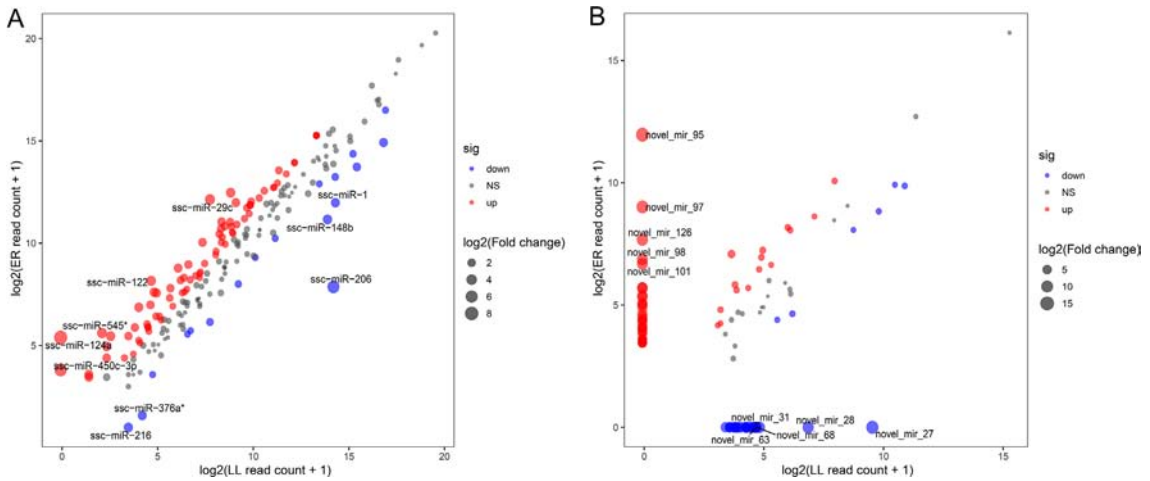


FIGURE 4 | Differential expression of porcine known **(A)** and novel **(B)** miRNAs between ER and LL. Each point in the figure represents the $\log_2$(ER/LL read count + 1) of an miRNA. Red points represent miRNAs with $\log_2$(ER/LL) > 1 and adjusted $P < 0.05$, blue points represent miRNAs with $\log_2$(ER/LL) < −1 and adjusted $P < 0.05$, and green points represent miRNAs with 1 > $\log_2$(ER/LL) > −1. The size of points shows the value of $\log_2$(ER/LL).

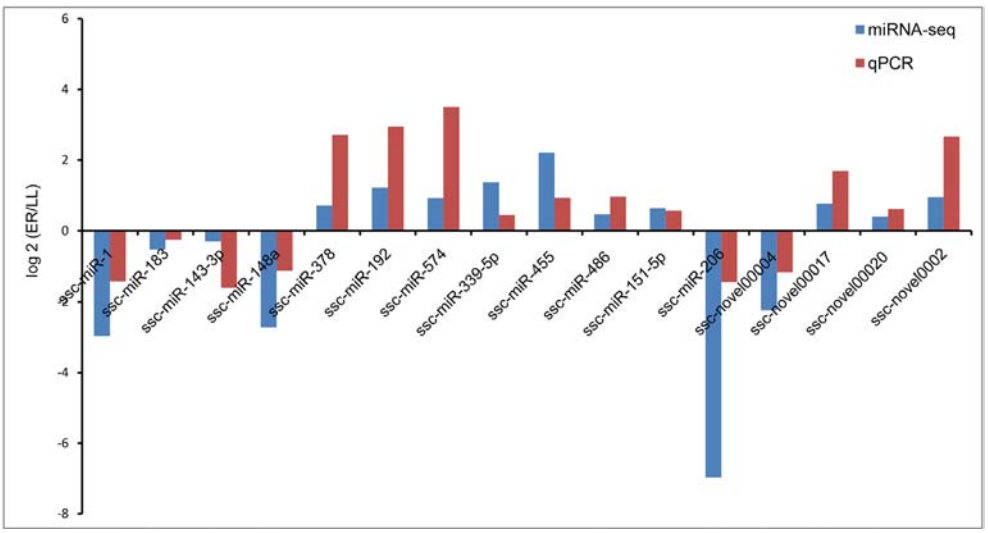


FIGURE 5 | Validation of deep sequencing results *via* real-time qRT-PCR.

miR-206 Directly Targeted 3′-UTR of IGF-1 and Inhibited its Protein Expression

miR-206 was the top DEM, and it was expressed in both LL and ER pigs. To uncover the function of miR-206, the potential targets of miR-206 in the data bank (www.targetscan.org; **Figure 7A**) was examined, and IGF-1 was identified as a prime target, with a highly conserved complementary miR-206-binding site in its 3′-UTR across mammals from humans to pandas (**Figure 7B**). As shown in **Figure 7**, the dual-luciferase reporter assay system showed a significant reduction of the *Renilla*/firefly luciferase ratio in the wild-type miRNA mimic compared to the mutant-type group. However, miR-206 had no appreciable inhibitory effect on a mutated IGF-1 3′-UTR dual-luciferase construct (**Figures 7C, D**). These results demonstrate the specific inhibition of IGF-1 expression by miR-206. miRNA regulates gene expression at the transcriptional level or at translational levels. To determine the regulation mechanism of miR-206, qRT-PCR and Western blot analysis were performed. Although no significant inhibition was detected at the IGF-1 mRNA level in porcine skeletal muscle SCs that were infected with pri-miR-206 expression lentivirus (**Figures 7E, F**), the inhibitory effect of miR-206 on IGF-1 protein expression was determined by Western blot analysis (**Figure 7G**). Therefore, it was confirmed

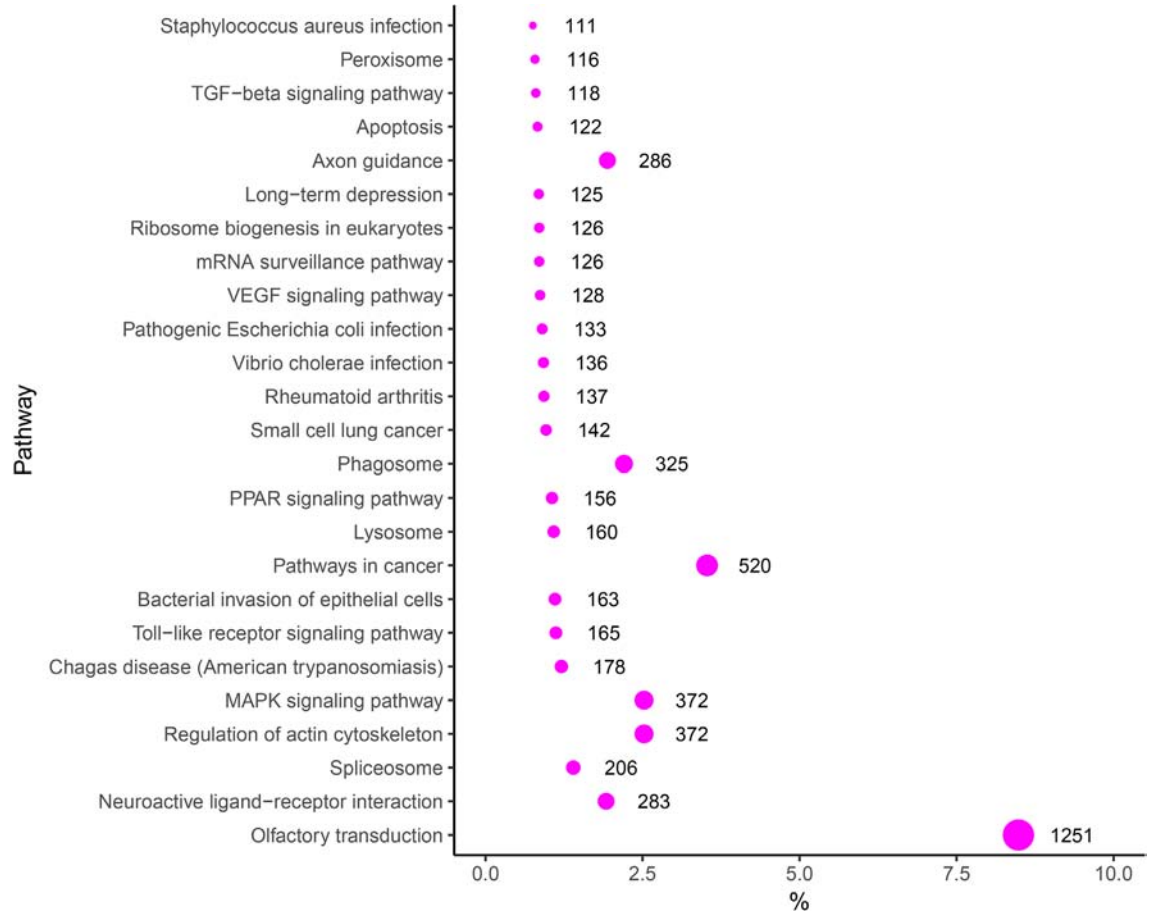


FIGURE 6 | KEGG enrichment scatter plot of the targets of DEMs.

that miR-206 directly targeted 3′-UTR of IGF-1 and inhibited its protein expression but not its mRNA transcription.

DISCUSSION

The miRNA system is a huge regulatory network of cellular processes, with a single miRNA being able to post-transcriptionally silence multiple mRNAs, while each mRNA can be targeted by numerous miRNAs (Friedman et al., 2008; Lu and Clark, 2012). In humans, more than 30% of the mRNAs are predicted to be miRNA targets (Griffiths-Jones et al., 2007). Recently, some miRNAs were found to be associated with endometrial receptivity, embryo development, and implantation (Ariel et al., 2011; Liu et al., 2016). In the present study, the sRNA profiles of endometrial tissues from ER and LL swine endometrium on GD12 using sequencing technology were compared to understand the miRNA-mediated regulation of embryo implantation. Our studies revealed the differential expression of 96 known miRNAs and 68 novel miRNAs in ER and LL endometrium, and the identification of miRNAs and target genes may be useful to develop new techniques and strategies for improving embryonic survival during implantation.

Highly Abundant miRNAs Might Affect Endometrial Remodeling

ssc-miR-143-3p, ssc-let-7a, and ssc-miR-21 were the top three miRNAs that were highly expressed in both LL and ER libraries. They were also found to be highly expressed in the endometrium of Meishan and Yorkshire pigs during early gestation in a recent published paper (Li et al., 2018). Mu et al. found that miR-143-3p inhibits proliferation and induced apoptosis in human hypertrophic scar fibroblast cells, and it also inhibited extracellular matrix production-associated protein expression (Mu et al., 2016). Several other studies also demonstrated that miR-143-3p suppressed proliferation and induced apoptosis in different carcinoma cells (He et al., 2016; Chen et al., 2017); thus, the highly expressed miR-143-3p in the porcine endometrium might also play a role in regulating the proliferation and apoptosis of endometrial cells. For let-7a, a functional investigation also revealed that it suppressed the proliferation of endometrial carcinoma (Liu et al., 2013b), and another study demonstrated that it markedly suppressed the proliferation, migration, and invasion of gastric cancer cells by down-regulating PKM2 (Tang et al., 2016). Furthermore, let-7a is involved in regulating the implantation process by the modulation of the expression of integrin-β3 and mucin 1 (Liu et al., 2012; Inyawilert et al., 2015).

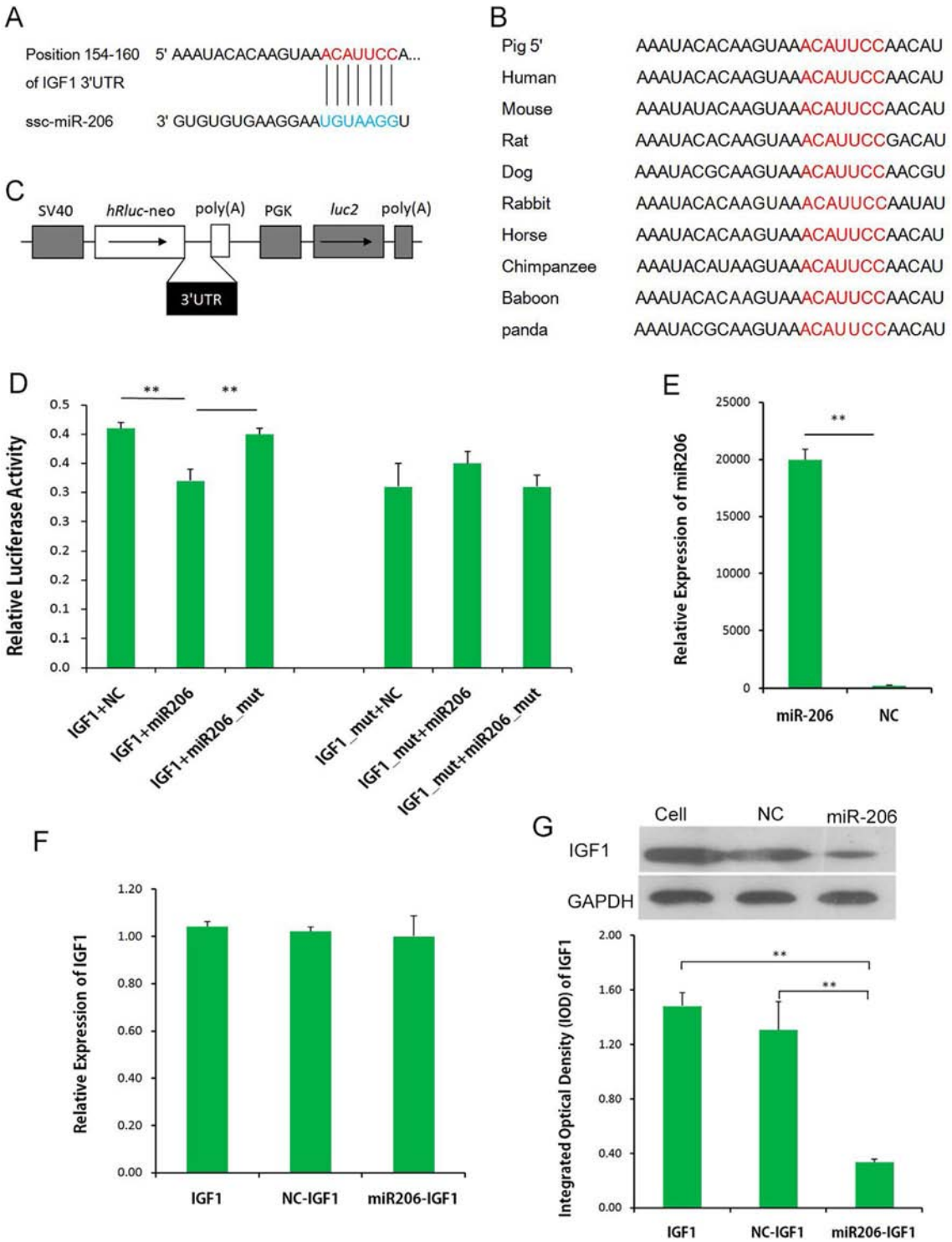


FIGURE 7 | miR-206 targets the 3′-UTR of IGF-1. **(A)** Predicted binding site of miR-206 in the 3′-UTR of IGF-1. **(B)** Binding site of miR-206 is highly conserved among mammals. **(C)** IGF-1 3′-UTR was inserted into the pmirGLO dual-luciferase reporter vector at the 3′-end of the *Renilla* luciferase gene (hRluc). **(D)** IGF-1 3′-UTR or IGF-1 3′-UTR-Mutant construct was co-transfected with miR-206, miR-206_mut, or NC, as indicated, into PK cells, and normalized *Renilla* luciferase activity was determined. **(E)** Expression of miR-206 in the porcine skeletal muscle SCs infected with pri-miR-206 expression lentivirus or NC lentivirus. **(F)** Expression of IGF-1 mRNA in SCs and SCs infected with pri-miR-206 expression lentivirus or NC lentivirus. **(G)** Expression of IGF-1 protein in SCs and SCs infected with pri-miR-206 expression lentivirus or NC lentivirus. Results are mean ± SD (three independent replicates per group). $^{*}P < 0.05$; $^{**}P < 0.01$ (Student's *t* test).

In addition, miR-21 has been causally linked to cellular proliferation, apoptosis, and migration in a wide variety of cancers (Asangani et al., 2008; Frankel et al., 2008). Previous studies have suggested that miR-21 was involved in embryo implantation in mouse (Hu et al., 2008), and a recent study provided evidence that miR-21 expressed in extracellular vesicles is very important in preimplantation embryo development (Lv et al., 2018). The KEGG pathway analysis of common miRNAs indicated that the predicted target genes were enriched in 1) cell self-renewal and degradation, including "Apoptosis," "Autophagy," "Ubiquitin-mediated proteolysis," "Longevity-regulating pathway," and "AMPK signaling pathway"; 2) cell motility, including "Regulation of actin cytoskeleton," "Focal adhesion," "ECM-receptor interaction," and "Rap1 signaling pathway"; and 3) cell proliferation and differentiation, including "FoxO signaling pathway," "mTOR signaling pathway," and "MAPK signaling pathway," Based on these results, it can be inferred that most highly abundant miRNAs in porcine endometrium mainly played important roles in regulating endometrial remodeling at the time of implantation.

Differentially Expressed Known miRNAs Related to Proliferation and Angiogenesis

Of the 96 differentially expressed known miRNAs, more than 80% were up-regulated in ER compared to LL sows; among them, ssc-miR-29c was the top miRNA that was expressed in both breeds (**Supplementary Table S7**). A recent study demonstrated that miR-29c affects human endometrial cells by suppressing cell proliferation and invasion as well as promotes cell apoptosis by inhibiting c-Jun expression (Long et al., 2015). It was also shown previously that miR-29c inhibited cell proliferation and induced apoptosis in many types of carcinoma cells (Wang et al., 2011; Liu et al., 2013a). Similarly, ssc-miR-214, which was among the top five miRNAs up-regulated in ER compared to LL sows (**Supplementary Table S7**), also played a role in promoting apoptosis and suppressing cell proliferation in several types of cells (Feng et al., 2011; Yang et al., 2013; Zhang et al., 2014). These results indicate that the markedly higher expression of miRNAs in ER than in LL (i.e., miR-29c and miR-214) could also play a key role in inhibiting endometrial cell proliferation and invasion, which could contribute to developing a more stable uterine environment. The results are also consistent with the authors' previous findings, which also revealed that the endometrium of ER pigs had a lower growth-promoting ability (Zhang et al., 2013). Strong evidence has shown that increased prolificacy of Chinese Taihu pigs might be due to an increased embryonic survival resulting from the more stable uterine environment and increased uterine receptivity (Stroband et al., 1992; Youngs et al., 1993; Youngs et al., 1994).

Pathway analyses can provide a better understanding of the molecular functions and biological processes of target genes. Among the target genes of differentially expressed known miRNAs, some KEGG pathways that are important for reproduction were significantly enriched. Notably, the mitogen-activated protein kinase (MAPK) signaling pathway, the Toll-like receptor signaling pathway, the peroxisome proliferator-activated receptor (PPAR) signaling pathway, the vascular endothelial growth factor (VEGF) signaling pathway, and the transforming growth factor-β (TGF-β) signaling pathway were in the top 25 signaling pathways. The MAPK signaling pathway is involved in the regulation of human endometrial cell proliferation (Park et al., 2017; Zhang et al., 2018). In another study, research data indicated that the activation of the MAPK signaling pathway can increase the proliferation of porcine uterine LE cells and may affect implantation in early pregnancy in pigs (Lim et al., 2018). In addition, the PPAR and TGF-β signaling pathways were also related to cell proliferation and had influence on implantation (Lim and Dey, 2000; Li et al., 2004; Chang et al., 2008; Tsang et al., 2013; Cheng et al., 2017; Song et al., 2018). The Toll-like receptor signaling pathway is responsible for innate immune responses, and studies provide evidence that this pathway takes part in implantation by regulating trophoblast cells' adhesion to endometrial cells (Montazeri et al., 2016). Studies provide evidence that the VEGF signaling pathway is known as the regulator of several endothelial cell functions, including mitogenesis, permeability, vascular tone, and the production of vasoactive molecules (Giles, 2001). Previous studies also indicated that it plays important roles in implantation and maintenance of pregnancy (Das et al., 1997; Halder et al., 2000; Möller et al., 2001; Hannan et al., 2011). Collectively, these pathway analyses illustrate some of the possible roles of highly expressed miRNAs in reproduction.

IGF-1 can regulate endothelial cell migration and promote angiogenesis (Shigematsu et al., 1999). In the human endometrium, it was found that IGF-1 participates in the maintenance of an angiogenic phenotype by inducing VEGF expression (Bermont et al., 2000). Furthermore, a recent study reported that IGF-1 is a critical determinant of neonatal porcine uterine development (George et al., 2018). Our results demonstrated that IGF-1 protein expression was directly inhibited by miR-206, which were highly expressed in LL and lowly expressed in ER. This suggests that the low expression of miR-206 in ER might facilitate the angiogenesis of endometrium during peri-implantation, but further studies are required to verify this hypothesis.

CONCLUSIONS

In summary, Illumina sequencing was used to identify 288 distinct miRNAs, consisting of 202 previously reported and 86 novel miRNAs, from porcine endometrium in two different reproduction capacity breeds. In a comparison of ER to LL sows, 96 significantly differentially expressed known miRNAs (78 up-regulated and 18 down-regulated) were identified. The target gene expression and pathway enrichment analyses indicated that these DEMs may influence embryonic implantation by regulating pathways related to proliferation, immunization, and angiogenesis. Our findings help gain a better understanding of the role of miRNAs in the regulation of embryonic implantation and embryonic survival in pigs. Future studies to identify target mRNAs regulated by abundant miRNAs in the endometrium using a single type of endometrial cell (i.e., luminal or glandular epitheliums) will be critical to uncover their exact biological functions.

DATA AVAILABILITY

The datasets used and analysed during the current study are available from the corresponding author on reasonable request. The raw reads produced in this study were deposited in the NCBI Sequence Read Archive (SRA), available using accession number PRJNA50573.

ETHICS STATEMENT

All researches involving animals were conducted according to animal ethics guidelines and approved by the Animal Care and Use Committee of South China Agricultural University (Guangzhou, China).

AUTHOR CONTRIBUTIONS

JL, ZW, and HZ designed the study. LH, RL, XQ, and SW performed the experiments, analyzed the data, and drafted the manuscript. XW performed the sequencing analysis. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31802033), the Guangdong Provincial Promotion Project on Preservation and UtiIization of Local Breed of Livestock and Poultry (4300-F18260), and the Science & Technology Planning Project of Guangzhou in China (201904010434). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank all the study participants, research staff, and students who assisted in animal sampling and technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00661/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | ClueGo network of pathways. Each node represents a pathway. The enrichment significance of pathway is reflected by the size of the nodes. Node color, represents the class that they belong. Mixed coloring means that the specific node belongs to multiple classes.

SUPPLEMENTARY TABLE 1 | Primer sequences for qPCR.

SUPPLEMENTARY TABLE 2 |The detailed sequence information in ER and LL small RNA libraries.

SUPPLEMENTARY TABLE 3 | Match_hairpin in ER library.

SUPPLEMENTARY TABLE 4 | Match_hairpin in LL library.

SUPPLEMENTARY TABLE 5 | Sequence editing of bases in the seed region of the miRNAsin ER and LL.

SUPPLEMENTARY TABLE 6 | Global expression profile of endometrial miRNAs on day 12 of gestation in LL and ER pigs.

SUPPLEMENTARY TABLE 7 | The differentially expressed miRNAs in ER and LL small RNA libraries.

REFERENCES

Altmäe, S., Martinez-Conejero, J. A., Esteban, F. J., Ruiz-Alonso, M., Stavreus-Evers, A., Horcajadas, J. A., et al. (2013). microRNAs miR-30b, miR-30d, and miR-494 regulate human endometrial receptivity. *Reprod. Sci.* 20 (3), 308–317. doi: 10.1177/1933719112453507

Aravin, A., and Tuschl, T. (2005). Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett.* 579 (26), 5830–5840. doi: 10.1016/j.febslet.2005.08.009

Ariel, R., Hanna, A., Juliet, S., Yoav, S., and Reuven, R. (2011). MicroRNAs are associated with human embryo implantation defects. *Human Reprod.* 26 (10), 2830–2840. doi: 10.1093/humrep/der255

Asangani, I. A., Rasheed, S. A., Nikolova, D., Leupold, J., Colburn, N., Post, S., et al. (2008). MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* 27 (15), 2128. doi: 10.1038/sj.onc.1210856

Ashworth, C., Pickard, A., Miller, S., Flint, A., and Diehl, J. (1997). Comparative studies of conceptus-endometrial interactions in Large White×Landrace and Meishan gilts. *Reprod. Fertil. Dev.* 9 (2), 217–226. doi: 10.1071/R96040

Audic, S., and Claverie, J.-M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7 (10), 986–995. doi: 10.1101/gr.7.10.986

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116 (2), 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bazer, F., Thatcher, W., Martinat-Botte, F., and Terqui, M. (1988). Conceptus development in Large White and prolific Chinese Meishan pigs. *J. Reprod. Fertil.* 84 (1), 37–42. doi: 10.1530/jrf.0.0840037

Bazer, F. W., and Johnson, G. A. (2014). Pig blastocyst-uterine interactions. *Differentiation* 87 (1), 52–65. doi: 10.1016/j.diff.2013.11.005

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57 (1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bermont, L., Lamielle, F., Fauconnet, S., Esumi, H., Weisz, A., and Adessi, G. L. (2000). Regulation of vascular endothelial growth factor expression by insulin-like growth factor-I in endometrial adenocarcinoma cells. *Int. J. Cancer* 85 (1), 117–123. doi: 10.1002/(SICI)1097-0215(20000101)85:1<117::AID-IJC21>3.0.CO;2-X

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., et al. (2004). GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20 (18), 3710–3715. doi: 10.1093/bioinformatics/bth456

Carrington, J. C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science* 301 (5631), 336–338. doi: 10.1126/science.1085242

Chang, H. J., Lee, J. H., Hwang, K. J., Kim, M. R., Chang, K. H., Park, D. W., et al. (2008). Transforming growth factor (TGF)-β1-induced human endometrial stromal cell decidualization through extracellular signal-regulated kinase and Smad activation in vitro: peroxisome proliferator-activated receptor gamma acts as a negative regulator of TGF-β1. *Fertil Steril* 90 (4), 1357–1365. doi: 10.1016/j.fertnstert.2007.09.010

Chen, L., Yao, H., Wang, K., and Liu, X. (2017). Long non-coding RNA MALAT1 regulates ZEB1 expression by sponging miR-143-3p and promotes hepatocellular carcinoma progression. *J. Cell. Biochem.* 118 (12), 4836–4843. doi: 10.1002/jcb.26158

Cheng, J.-C., Chang, H.-M., and Leung, P. C. K. (2017). TGF-β 1 inhibits human trophoblast cell invasion by upregulating connective tissue growth factor expression. *Endocrinology* 158 (10), 3620–3628. doi: 10.1210/en.2017-00536

Christenson, R., Vallet, J., Leymaster, K., and Young, L. (1993). Uterine function in Meishan pigs. *J. Reprod. Fertil. Suppl.* 48, 279–289.

Das, S., Chakraborty, I., Wang, J., Dey, S., and Hoffman, L. (1997). Expression of vascular endothelial growth factor (VEGF) and VEGF-receptor messenger ribonucleic acids in the peri-implantation rabbit uterus. *Biol. Reprod.* 56 (6), 1390–1399. doi: 10.1095/biolreprod56.6.1390

Feng, Y., Cao, J. H., Li, X. Y., and Zhao, S. H. (2011). Inhibition of miR-214 expression represses proliferation and differentiation of C2C12 myoblasts. *Cell Biochem. Funct.* 29 (5), 378–383. doi: 10.1002/cbf.1760

Frankel, L. B., Christoffersen, N. R., Jacobsen, A., Lindow, M., Krogh, A., and Lund, A. H. (2008). Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J. Biol. Chem.* 283 (2), 1026–1033. doi: 10.1074/jbc.M707224200

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40 (1), 37–52. doi: 10.1093/nar/gkr688

Friedman, R., Farh, K., Burge C., and Bartel, D. (2008). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19 (1), 92–105. doi: 10.1101/gr.082701.108

Geisert, R., Renegar, R., Thatcher, W., Roberts, R., and Bazer, F. (1982). Establishment of pregnancy in the pig: i. Interrelationships between preimplantation development of the pig blastocyst and uterine endometrial secretions. *Biol. Reprod.* 27 (4), 925–939. doi: 10.1095/biolreprod27.4.925

George, A. F., Rahman, K. M., Miller, D. J., Wiley, A. A., Camp, M. E., Bartol, F., et al. (2018). Effects of colostrum, feeding method, and oral IGF1 on porcine uterine development. *Reproduction* 155 (3), 259–271. doi: 10.1530/REP-17-0658

Giles, F. J. (2001). The vascular endothelial growth factor (VEGF) signaling pathway: a therapeutic target in patients with hematologic malignancies. *Oncologist* 6 (Supplement 5), 32–39. doi: 10.1634/theoncologist.6-suppl_5-32

Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2007). miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36 (suppl_1), D154–D158. doi: 10.1093/nar/gkm952

Halder, J., Zhao, X., Soker, S., Paria, B., Klagsbrun, M., Das, S., et al. (2000). Differential expression of VEGF isoforms and VEGF164-specific receptor neuropilin-1 in the mouse uterus suggests a role for VEGF164 in vascular permeability and angiogenesis during implantation. *Genesis* 26 (3), 213–224. doi: 10.1002/(SICI)1526-968X(200003)26:3<213::AID-GENE7>3.0.CO;2-M

Haley, C., Lee, G., and Ritchie, M. (1995). Comparative reproductive performance in Meishan and Large White pigs and their crosses. *Anim. Genet.* 60 (2), 259–267. doi: 10.1017/S1357729800008420

Hannan, N., Paiva, P., Meehan, K., Rombauts, L., Gardner, D., and Salamonsen, L. (2011). Analysis of fertility-related soluble mediators in human uterine fluid identifies VEGF as a key regulator of embryo implantation. *Endocrinology* 152 (12), 4948–4956. doi: 10.1210/en.2011-1248

He, Z., Yi, J., Liu, X., Chen, J., Han, S., Jin, L., et al. (2016). MiR-143-3p functions as a tumor suppressor by regulating cell proliferation, invasion and epithelial-mesenchymal transition by targeting QKI-5 in esophageal squamous cell carcinoma. *Mol. Cancer* 15 (1), 51. doi: 10.1186/s12943-016-0533-3

Hu, S.-J., Ren, G., Liu, J.-L., Zhao, Z.-A., Yu, Y.-S., Su, R.-W., et al. (2008). MicroRNA expression and regulation in mouse uterus during embryo implantation. *J. Biol. Chem.* 283 (34), 23473–23484. doi: 10.1074/jbc.M800406200

Inyawilert, W., Fu, T.-Y., Lin, C.-T., and Tang, P.-C. (2015). Let-7-mediated suppression of mucin 1 expression in the mouse uterus during embryo implantation. *J. Reprod. Dev.* 61 (2), 138–144. doi: 10.1262/jrd.2014-106

Jazdzewski, K., Liyanarachchi, S., Swierniak, M., Pachucki, J., Ringel, M. D., Jarzab, B., et al. (2009). Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc. Natl. Acad. Sci. U.S.A.* 106 (5), 1502–1505. doi: 10.1073/pnas.0812591106

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., and Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315 (5815), 1137–1140. doi: 10.1126/science.1138050

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25

Laurent, L. (2010). MicroRNAs in embryonic stem cells and early embryonic development. *J. Cell. Mol. Med.* 12 (6A), 2181–2188. doi: 10.1111/j.1582-4934.2008.00513.x

Li, G., Li, Y., Li, X., Ning, X., Li, M., and Yang, G. (2011). MicroRNA identity and abundance in developing swine adipose tissue as determined by Solexa sequencing. *J. Cell. Biochem.* 112 (5), 1318–1328. doi: 10.1002/jcb.23045

Li, Q., Cheon, Y.-P., Kannan, A., Shanker, S., Bagchi, I. C., and Bagchi, M. K. (2004). A novel pathway involving progesterone receptor, 12/15-lipoxygenase-derived eicosanoids, and peroxisome proliferator-activated receptor γ regulates implantation in mice. *J. Biol. Chem.* 279 (12), 11570–11581. doi: 10.1074/jbc.M311773200

Li, M., Xia, Y., Gu, Y., Zhang, K., Lang, Q., Chen, L., et al. (2010). MicroRNAome of porcine pre-and postnatal development. *PLoS One* 5 (7), e11541. doi: 10.1371/journal.pone.0011541

Li, W., Xi, Y., Xue, S., Wang, Y., Wu, L., Liu, H., et al. (2018). Sequence analysis of microRNAs during pre-implantation between Meishan and Yorkshire pigs. *Gene* 646, 20–27. doi: 10.1016/j.gene.2017.12.046

Lim, H., and Dey, S. K. (2000). PPARδ functions as a prostacyclin receptor in blastocyst implantation. *Trends Endocrinol. Metab.* 11 (4), 137–142. doi: 10.1016/S1043-2760(00)00243-5

Lim, W., Bae, H., Bazer, F. W., and Song, G. (2018). Characterization of CC motif chemokine ligand 4 in the porcine endometrium during the presence of the maternal-fetal interface. *Dev. Biol.* 441 (1), 146–158. doi: 10.1016/j.ydbio.2018.06.022

Liu, R., Wang, M., Su, L., Li, X., Zhao, S., and Yu, M. (2015). The expression pattern of microRNAs and the associated pathways involved in the development of porcine placental folds that contribute to the expansion of the exchange surface area. *Biol. Reprod.* 93 (3), 1–13. doi: 10.1095/biolreprod.114.126540

Liu, W., Niu, Z., Li, Q., Pang, R. T., Chiu, P. C., and Yeung, W. S. (2016). MicroRNA and Embryo Implantation. *Am. J. Reprod. Immunol.* 75 (3), 263–271. doi: 10.1111/aji.12470

Liu, N., Okamura, K., Tyler, D. M., Phillips, M. D., Chung, W.-J., and Lai, E. C. (2008). The evolution and functional diversification of animal microRNA genes. *Cell Res.* 18 (10), 985. doi: 10.1038/cr.2008.278

Liu, W.-M., Pang, R. T., Cheong, A. W., Ng, E. H., Lao, K., Lee, K.-F., et al. (2012). Involvement of microRNA lethal-7a in the regulation of embryo implantation in mice. *PLoS One* 7 (5), e37039. doi: 10.1371/journal.pone.0037039

Liu, N., Tang, L.-L., Sun, Y., Cui, R.-X., Wang, H.-Y., Huang, B.-J., et al. (2013a). MiR-29c suppresses invasion and metastasis by targeting TIAM1 in nasopharyngeal carcinoma. *Cancer Lett.* 329 (2), 181–188. doi: 10.1016/j.canlet.2012.10.032

Liu, P., Qi, M., Ma, C., Lao, G., Liu, Y., Liu, Y., et al. (2013b). Let7a inhibits the growth of endometrial carcinoma cells by targeting Aurora-B. *FEBS Lett.* 587 (16), 2523–2529. doi: 10.1016/j.febslet.2013.05.065

Long, M., Wan, X., La, X., Gong, X., and Cai, X. (2015). miR-29c is downregulated in the ectopic endometrium and exerts its effects on endometrial cell proliferation, apoptosis and invasion by targeting c-Jun. *Int. J. Mol. Med.* 35 (4), 1119–1125. doi: 10.3892/ijmm.2015.2082

Lu, J., and Clark, A. G. (2012). Impact of microRNA regulation on variation in human gene expression. *Genome Res.* 22 (7), 1243–1254. doi: 10.1101/gr.132514.111

Lv, C., Yu, W.-X., Wang, Y., Yi, D.-J., Zeng, M.-H., and Xiao, H.-M. (2018). MicroRNA-21 in extracellular vesicles contributes to the growth of fertilized eggs and embryo development in mice. *Biosci. Rep.* 38 (4), BSR20180036. doi: 10.1042/BSR20180036

Möller, B., Rasmussen, C., Lindblom, B., and Olovsson, M. (2001). Expression of the angiogenic growth factors VEGF, FGF-2, EGF and their receptors in normal human endometrium during the menstrual cycle. *Mol. Hum. Reprod.* 7 (1), 65–72. doi: 10.1093/molehr/7.1.65

Montazeri, M., Sanchez-Lopez, J., Caballero, I., Maslehat Lay, N., Elliott, S., and Fazeli, A. (2016). Interleukin-1 receptor antagonist mediates toll-like receptor 3-induced inhibition of trophoblast adhesion to endometrial cells in vitro. *Human Reprod.* 31 (9), 2098–2107. doi: 10.1093/humrep/dew171

Montenegro, D., Romero, R., Kim, S.-S., Tarca, A. L., Draghici, S., Kusanovic, J. P., et al. (2009). Expression patterns of microRNAs in the chorioamniotic membranes: a role for microRNAs in human pregnancy and parturition. *J. Pathol. Bacteriol.* 217 (1), 113–121. doi: 10.1002/path.2463

Mu, S., Kang, B., Zeng, W., Sun, Y., and Yang, F. (2016). MicroRNA-143-3p inhibits hyperplastic scar formation by targeting connective tissue growth factor CTGF/CCN2 via the Akt/mTOR pathway. *Mol. Cell. Biochem.* 416 (1–2), 99–108. doi: 10.1007/s11010-016-2699-9

Nicoli, S., Standley, C., Walker, P., Hurlstone, A., Fogarty, K. E., and Lawson, N. D. (2012). MicroRNA-mediated integration of haemodynamics and VEGF signalling during angiogenesis. *Nature* 464 (7292), 1196. doi: 10.1038/nature08889

Nielsen, M., Hansen, J., Hedegaard, J., Nielsen, R., Panitz, F., Bendixen, C., et al. (2010). MicroRNA identity and abundance in porcine skeletal muscles determined by deep sequencing. *Anim. Genet.* 41 (2), 159–168. doi: 10.1111/j.1365-2052.2009.01981.x

Park, S., Lim, W., Bazer, F. W., and Song, G. (2017). Naringenin induces mitochondria-mediated apoptosis and endoplasmic reticulum stress by regulating MAPK and AKT signal transduction pathways in endometriosis cells. *MHR: Basic Sci. Reprod. Med.* 23 (12), 842–854. doi: 10.1093/molehr/gax057

Pope, W. F., and First, N. L. (1985). Factors affecting the survival of pig embryos. *Theriogenology* 23 (1), 91–105. doi: 10.1016/0093-691X(85)90075-5

Qian, K., Hu, L., Chen, H., Li, H., Liu, N., Li, Y., et al. (2009). Hsa-miR-222 is involved in differentiation of endometrial stromal cells in vitro. *Endocrinology* 150 (10), 4734–4743. doi: 10.1210/en.2008-1629

Rosenbluth, E. M., Shelton, D. N., Sparks, A. E. T., Eric, D., Lane, C., and Van Voorhis, B. J. (2013). MicroRNA expression in the human blastocyst. *Fertil. Steril.* 99 (3), 855–861.e853. doi: 10.1016/j.fertnstert.2012.11.001

Scofield, A. M., Clegg, F. G., and Lamming, G. E. (1974). Embryonic mortality and uterine infection in the pig. *J. Reprod. Fertil.* 36 (2), 353–361. doi: 10.1530/jrf.0.0360353

Shigematsu, S., Yamauchi, K., Nakajima, K., Iijima, S., Aizawa, T., and Hashizume, K. (1999). IGF-1 regulates migration and angiogenesis of human endothelial cells. *Endocr. J.* 46 (Suppl), S59–S62. doi: 10.1507/endocrj.46.Suppl_S59

Song, T., Zhao, J., Jiang, T., Jin, X., Li, Y., and Liu, X. (2018). Formononetin protects against balloon injury-induced neointima formation in rats by regulating proliferation and migration of vascular smooth muscle cells via the TGF-β1/Smad3 signaling pathway. *Int. J. Mol. Med.* 42 (4), 2155–2162. doi: 10.3892/ijmm.2018.3784

Spencer, T. E., Johnson, G. A., Bazer, F. W., Burghardt, R. C., and Palmarini, M. (2006). Pregnancy recognition and conceptus implantation in domestic ruminants: roles of progesterone, interferons and endogenous retroviruses. *Reprod. Fertil. Dev.* 19 (1), 65–78. doi: 10.1071/RD06102

Stroband, H., Voolstra, C., Taverne, N., and Hazeleger, W. (1992). Morphological indications of secretion by uterine epithelial cells and changes in the ratio between acidic and basic uterine proteins during early pregnancy in Chinese Meishan pig. *Eur J Morphol* 30 (3), 205–218.

Su, L., Zhao, S., Zhu, M., and Yu, M. (2010). Differential expression of microRNAs in porcine placentas on days 30 and 90 of gestation. *Reprod. Fertil. Dev.* 22 (8), 1175–1182. doi: 10.1071/RD10046

Tang, R., Yang, C., Ma, X., Wang, Y., Luo, D., Huang, C., et al. (2016). MiR-let-7a inhibits cell proliferation, migration, and invasion by down-regulating PKM2 in gastric cancer. *Oncotarget* 7 (5), 5972. doi: 10.18632/oncotarget.6821

Tsang, H., Cheung, T.-Y., Kodithuwakku, S. P., Chai, J., Yeung, W. S., Wong, C. K., et al. (2013). Perfluorooctanoate suppresses spheroid attachment on endometrial epithelial cells through peroxisome proliferator-activated receptor alpha and down-regulation of Wnt signaling. *Reprod. Toxicol.* 42, 164–171. doi: 10.1016/j.reprotox.2013.08.001

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131 (4), 281–285. doi: 10.1007/s12064-012-0162-3

Wang, C.-M., Wang, Y., Fan, C.-G., Xu, F.-F., Sun, W.-S., Liu, Y.-G., et al. (2011). miR-29c targets TNFAIP3, inhibits cell proliferation and induces apoptosis in hepatitis B virus-related hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* 411 (3), 586–592. doi: 10.1016/j.bbrc.2011.06.191

Wen, M., Shen, Y., Shi, S., and Tang, T. (2012). miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. *BMC Bioinformatics* 13 (1), 140. doi: 10.1186/1471-2105-13-140

Wessels, J. M., Edwards, A. K., Khalaj, K., Kridli, R. T., Bidarimath, M., and Tayade, C. (2013). The microRNAome of pregnancy: deciphering miRNA networks at the maternal-fetal interface. *PLoS One* 8 (11), e72264. doi: 10.1371/journal.pone.0072264

Wilson, M. E., Biensen, N. J., and Ford, S. P. (1999). Novel insight into the control of litter size in pigs, using placental efficiency as a selection tool. *J. Anim. Sci.* 77 (7), 1654–1658. doi: 10.2527/1999.7771654x

Wu, L., Fan, J., and Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 103 (11), 4034–4039. doi: 10.1073/pnas.0510928103

Yang, T.-S., Yang, X.-H., Wang, X.-D., Wang, Y.-L., Zhou, B., and Song, Z.-S. (2013). MiR-214 regulate gastric cancer cell proliferation, migration and invasion by targeting PTEN. *Cancer Cell Int.* 13 (1), 68. doi: 10.1186/1475-2867-13-68

Youngs, C., Christenson, L., and Ford, S. (1994). Investigations into the control of litter size in swine: III. A reciprocal embryo transfer study of early conceptus development. *J. Anim. Sci.* 72 (3), 725–731. doi: 10.2527/1994.723725x

Youngs, C., Ford, S., McGinnis, L., and Anderson, L. (1993). Investigations into the control of litter size in swine: I. Comparative studies on in vitro development of Meishan and Yorkshire preimplantation embryos. *J. Anim. Sci.* 71 (6), 1561–1565. doi: 10.2527/1993.7161561x

Zavy, M. T., and Geisert, R. D. (1994). *Embryonic mortality in domestic species*. FL: CRC Press Boca Raton.

Zhang, Z. (1986). *Chinese pig breed records*. Shanghai, China: Shanghai Science and Technology Press.

Zhang, G., Cheng, Y., Zhang, Q., Li, X., Zhou, J., Wang, J., et al. (2018). ATX-LPA axis facilitates estrogen-induced endometrial cancer cell proliferation via MAPK/ERK signaling pathway. *Mol. Med. Rep.* 17 (3), 4245–4252. doi: 10.3892/mmr.2018.8392

Zhang, H., Wang, S., Liu, M., Zhang, A., Wu, Z., Zhang, Z., et al. (2013). Differential gene expression in the endometrium on gestation day 12 provides insight into sow prolificacy. *BMC Genomics* 14 (1), 45. doi: 10.1186/1471-2164-14-45

Zhang, Z.-C., Li, Y.-Y., Wang, H.-Y., Fu, S., Wang, X.-P., Zeng, M.-S., et al. (2014). Knockdown of miR-214 promotes apoptosis and inhibits cell proliferation in nasopharyngeal carcinoma. *PLoS One* 9 (1), e86149. doi: 10.1371/journal.pone.0086149

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hong, Liu, Qiao, Wang, Wang, Li, Wu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 August 2019
doi: 10.3389/fgene.2019.00691

Differentially Expressed MiRNAs and tRNA Genes Affect Host Homeostasis During Highly Pathogenic Porcine Reproductive and Respiratory Syndrome Virus Infections in Young Pigs

Damarius S. Fleming[1,2] *and Laura C. Miller*[2]*

[1] ORAU/ORISE, Oak Ridge, TN, United States, [2] Virus and Prion Diseases of Livestock Research Unit, National Animal Disease Center, USDA, Agricultural Research Service, Ames, IA, United States.

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin,
Ireland

Reviewed by:
Elisabetta Giuffra,
INRA Centre Jouy-en-Josas,
France
Carolina Neves Correia,
University College Dublin,
Ireland

***Correspondence:**
Laura C. Miller
laura.miller@ars.usda.gov

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 07 March 2019
Accepted: 02 July 2019
Published: 02 August 2019

Citation:
Fleming DS and Miller LC (2019) Differentially Expressed MiRNAs and tRNA Genes Affect Host Homeostasis During Highly Pathogenic Porcine Reproductive and Respiratory Syndrome Virus Infections in Young Pigs. Front. Genet. 10:691. doi: 10.3389/fgene.2019.00691

Background: Porcine respiratory and reproductive syndrome virus (PRRSV) is a single-stranded RNA virus member that infects pigs and causes losses to the commercial industry reaching upward of a billion dollars annually in combined direct and indirect costs. The virus can be separated into etiologies that contain multiple heterologous low and highly pathogenic strains. Recently, the United States has begun to see an increase in heterologous type 2 PRRSV strains of higher virulence (HP-PRRSV). The high pathogenicity of these strains can drastically alter host immune responses and the ability of the animal to maintain homeostasis. Because the loss of host homeostasis can denote underlying changes in gene and regulatory element expression profiles, the study aimed to examine the effect PRRSV infections has on miRNA and tRNA expression and the roles they play in host tolerance or susceptibility.

Results: Using transcriptomic analysis of whole blood taken from control and infected pigs at several time points (1, 3, 8 dpi), the analysis returned a total of 149 statistically significant (FDR ⩽ 0.15) miRNAs (n = 89) and tRNAs (n = 60) that were evaluated for possible pro- and anti-viral effects. The tRNA differential expression increased in both magnitude and count as dpi increased, with no statistically significant expression at 1 dpi, but increases at 3 and 8 dpi. The most abundant tRNA amino acid at 3 dpi was alanine, while glycine was the most abundant at 8 dpi. For the miRNAs, focus was put on upregulation that can inhibit gene expression. These results yielded candidates with potential anti- and pro-viral actions such as Ssc-miR-125b, which is predicted to limit PRRSV viral levels, and Ssc-miR-145-5p shown to cause alternative macrophage priming. The results also showed that both the tRNAs and miRNAs displayed expression patterns.

Conclusions: The results indicated that the HP-PRRSV infection affects host homeostasis through changes in miRNA and tRNA expression and their subsequent gene interactions that target and influence the function of host immune, metabolic, and structural pathways.

Keywords: miRNA—microRNA, tRNA, differential gene expression, porcine reproductive and respiratory syndrome virus, whole blood, pigs (*Sus scrofa*)

INTRODUCTION

Porcine respiratory and reproductive syndrome virus (PRRSV) is a single-stranded RNA virus member of the Nidovirales order that infects pigs and causes losses to the commercial swine industry that reach upward of a billion dollars annually in combined (Holtkamp et al., 2013) direct and indirect costs. The virus can be separated into etiologies that contain multiple heterologous strains due to the high mutation and recombination rates observed within the virus, leading to the evolution of both low and highly pathogenic PRRSV strains. Within industry herds, the low pathogenic strains can cause persistent infections that can last the entirety of the pigs "commercial life," while the highly pathogenic strains often cause acute illness and increased mortality. Although PRRSV infections can be found globally, a highly pathogenic Chinese type 2 PRRSV strain, not present in the United States, has ravaged much of Asia by presenting as an acute infection leading to high mortality in animals early in the commercial process. Additionally, recent studies have observed that the United States has begun to see an increase in heterologous type 2 PRRSV strains of increased virulence (van Geelen et al., 2018). The high pathogenicity of these strains can drastically alter host immune responses and the ability of the animal to maintain homeostasis. Because changes in the homeostasis of an individual can denote underlying changes in health and gene expression profiles, researchers have conducted studies to evaluate the host–virus interaction in the attempt to better understand the genetics involved in the host immune response to PRRSV (Shanmukhappa et al., 2007; Miller et al., 2012; Xiao et al., 2015; Miller et al., 2017).

The health of an animal, prior to and after illnesses, is, in many ways, a measure of the ability of that individual to maintain or return to a homeostatic state in which the interplay between immunologic and metabolic responses are in balance (Hotamisligil, 2006). In livestock, this balance is crucial to the health and growth of the animal. This change in internal balance of the host is most evident at the whole animal level, in which clinical signs of an illness can be represented by phenotypes such as fever, lameness, and changes in growth. However, these changes start at the genomic level where dysregulation can be represented as perturbations in the host ability to maintain proper communication between cellular receptors and signaling. This change in signaling cascades facilitates invasion of microbes, like PRRSV viral particles, into the host system as the miscommunication ameliorates pattern recognition receptor (PRR) evasion by the virus (Bowie and Unterholzner, 2008). The cell tropism of PRRSV, in general, is the monocyte-derived cells of the innate arm of immunity that become pulmonary alveolar macrophages in porcine lungs (Duan et al., 1997). During viral infections, host biological processes can become inundated from the activity of the viral life cycle, leading to enzymatic changes that alter host metabolic profiles. This in turn can lead to misfunction of macrophage activation signaling that indicates a dysregulated immunologic–metabolic axis (Hicks et al., 2013; Xiao et al., 2015; Langston et al., 2017).

It has been established that susceptibility to PRRSV infection and persistence involves a host genetic component (Dekkers et al., 2017), in which certain genes can behave in either a pro-viral or antiviral manner. Less understood are the actions of small noncoding regulatory and effector RNAs that influence host immunologic and metabolic functions to skew away from homeostasis during PRRSV infection. To begin to understand the effect of small noncoding RNAs (sncRNAs) on the host response to PRRSV, researchers have mostly examined the effects of miRNAs, centered around the canonical functions that miRNAs encompass through the inhibition of gene expression by transcriptional suppression of mRNA and the ability to be both pro- and antiviral (Bruscella et al., 2017). In order to identify and classify other host sncRNA classes affected by PRRSV infection, Fleming and Miller (2018) examined the landscape of porcine whole blood sncRNA. The study revealed whole blood to be a rich landscape of multiple sncRNA, in which to examine the regulatory changes within the host. The current study follows up on the results of Fleming and Miller (2018) using the same dataset with updates to the methods from the previous study to allow for transcriptomic analysis. The current study examined the differential expression of both miRNAs and tRNAs and the effect their expression has on host homeostasis during highly pathogenic type 2 PRRSV infections by examining the biological pathways in which they belong. The purpose and potential impact of the current study are to give insight into the regulatory functions of sncRNAs involved in host cellular communication and homeostasis and into how they relate to how host immune responses, like pro- and anti-inflammatory cytokine signaling, are tempered during infection with a highly pathogenic type 2 PRRSV strain.

METHODS

Animals and Sample Collection

PRRSV-free 3-week-old crossbreed pigs (Landrace × Yorkshire × Duroc) were purchased from a USDA-approved vendor (Wilson Farms, Wisconsin). Sample collection consisted of whole blood samples (~2.5 ml/pig) collected from twenty-eight 9-week-old anesthetized pigs by jugular venipuncture. The piglets were inoculated with either a sham inoculation (prepared from MARC-145 cell culture used to propagate the virus) for the controls (N = 14) (2 ml/pig) or challenged (N = 14) with an infectious cDNA clone of a Chinese highly pathogenic (HP) PRRSV strain isolate rJXwn06 (10^4 $TCID_{50}$/ml, 2 ml/pig). Whole blood sample collection occurred over several time points consisting of 1, 3, and 8 days post infection (dpi) using PAXgene® tubes. Blood samples were stored at -20 C prior to total RNA isolation and NGS library creation. At the conclusion of the study, the pigs were euthanized in the following manner: the animal was physically restrained for the intravenous administration of a barbiturate (Fatal Plus, Vortech Pharmaceuticals, Dearborn, MI) following the manufacturer label dose (1 ml/4.54 kg). Only 24 samples were used due to four animals succumbing prior to 8 dpi.

RNA Isolation and Sequencing Library Preparation

Total RNA was extracted from 2.5-ml cryopreserved whole blood samples using the protocol from Fleming and Miller (2018) and a modified miRNA extraction kit protocol

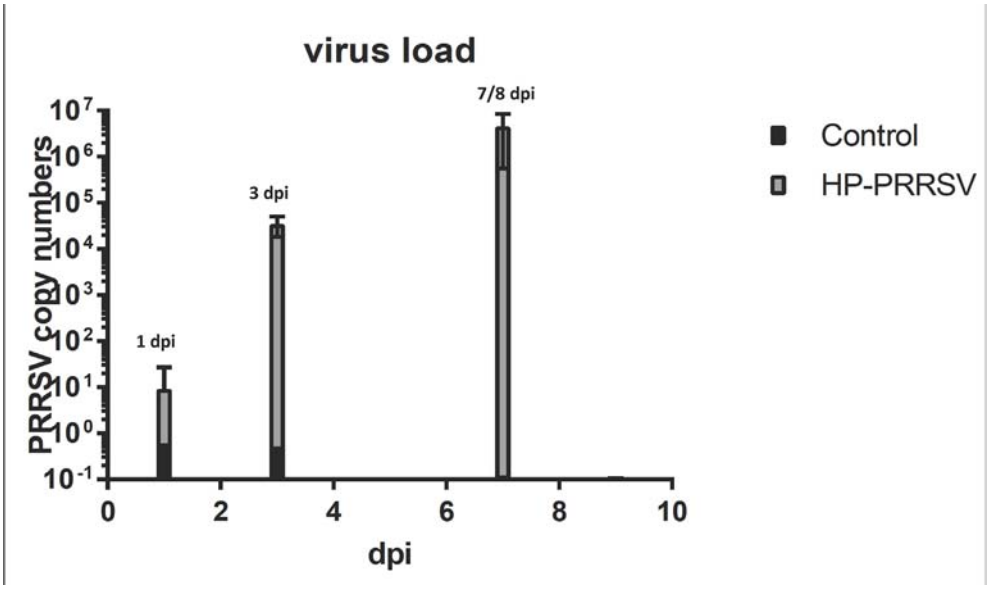


FIGURE 1 | Plot of viral load over time. The graph shows that HP-PRRSV replication increased across the course of infection. The largest increase takes place between 1 and 3 dpi that tracks with large changes in miRNA expression.

optimized according to Taxis et al. (2017) for the PAXgene® miRNA and MirVana miRNA isolation kit™ (Thermo Scientific, Wilmington, DE, USA). Optimization of these protocols was done to increase small RNA recovery for downstream library creation. All RNA was globin-depleted to account for high levels of globin transcripts using porcine-specific hemoglobin A and B (HBA and HBB) oligonucleotides based on the procedure from Choi et al. (2014). After extraction and globin reduction, the quality and concentration of the total RNA (N = 24) was checked by NanoDrop using each sample. Prior to library creation, sample quality was checked using the Agilent Bioanalyzer 1000 that showed the total RNA quality ranged from a RIN # of 6.5–9.2, and 260/280-nm concentrations ratios were at or above 2 for all samples after globin reduction prior to library preparation. Following quality control checks, the globin-reduced total RNA samples were then used for library preparation for sncRNA sequence generation. Library creation was carried out on 24 samples using the manufacturer's protocol for the NEBNext multiplex small RNA library prep kit® with a starting RNA amount of ~220 ng to ~1.1 µg. Step one of the protocol, the 3' SR adaptor ligation, was modified from the manufacturer's protocol to incubate the samples for 18 h at 16°C to increase ligation efficiency of methylated RNAs. Next, the 5' SR adaptor was ligated, and hybridization of the reverse transcription primer was performed. Samples were then barcoded using the NEBnext indexes (1–24) to allow for multiplexing prior to PCR amplification of cDNA. Samples were then put through a cleanup step using the Qiagen QIAquick PCR purification kit®, then checked for quality using the Agilent Bioanalyzer. Small RNA libraries were not size-selected to allow for the capture of multiple sncRNAs between 18–200 nt. Sequencing was carried out on the Illumina Hiseq 3000™ at the Iowa State University genomic sequencing center in Ames, IA to produce a total of one 100-bp single-end read for each of the 24 samples.

Transcriptomic Analysis

The transcriptomic analysis was accomplished using *in silico* resources within the Galaxy web interface (Blankenberg et al., 2010; Afgan et al., 2016). Quality assessment and control were performed using FastQC and TrimGalore (Martin, 2011) to remove adaptors and barcodes from multiplexing. Reads with a quality score below 38 and length less than 18 or longer than 72 nucleotides were discarded. A total of 24 sequences were generated for downstream analysis. The sequenced reads were mapped to the *S.scrofa* 10.2 reference genome from Ensembl using the Hisat2 (Kim et al., 2015) package with default settings. Annotation of gene counts was performed using FeatureCounts software package (Liao et al., 2014) coupled with an in-house-created sncRNA GTF file. The in-house GTF file was based on annotations from release 21 of miRbase, GtRNAdb using tRNAscan-SE 2.0, Ensembl 84 ncRNA database, and the RTH *S.scrofa* 10.2 ncRNA database. The differential gene expression was calculated using the DeSeq2 package with the dispersion model set as local with all other parameters set at their default values. The differential expression was based on the model ~Treatment $_{\text{HP-PRRSV, Control}}$ + $\text{Time}_{1,3,8}$ + Treatment : Time. All reported results are based on the interaction of treatment and time and were considered statistically significant at FDR ≤ 0.15 based on a Benjamini and Hochberg FDR adjustment. No cutoff was used for $\log_2$FC. Venn diagram analysis was conducted using the website http://bioinformatics.psb.ugent.be/webtools/Venn/ (2017). The 10.2 reference genome was used in lieu of the newer 11.1 version due to a current lack of updates to the miRNAs and tRNAs examined within the study. Data has been deposited in a public repository under GEO accession: GSE121980.

TABLE 1 | Ten most differentially expressed tRNAs and miRNAs by dpi. **(A)** tRNA. **(B)** miRNA.

(A) | Ten most differentially expressed tRNAs by dpi.

DPI	GeneID	tRNA ID	log_2(FC)
1	NONE	NONE	NONE
DPI	**GeneID**	**tRNA ID**	**log_2(FC)**
3	trna1668	ValAAC	2.07
3	trna1634	LeuCAA	1.52
3	trna660	TrpCCA	1.47
3	trna201	CysGCA	1.40
3	trna202	LeuTAG	1.28
3	trna1647	AlaCGC	−1.27
3	trna1245	ProTGG	−1.37
3	trna1635	AlaAGC	−1.58
3	trna1637	AlaCGC	−1.75
3	trna1503	ProAGG	−1.78
DPI	**GeneID**	**tRNA ID**	**log_2(FC)**
8	trna783	GlyGCC	2.37
8	trna630	GlyCCC	1.94
8	trna658	ProCGG	1.89
8	trna1245	ProTGG	1.86
8	trna1270	TyrGTA	1.84
8	trna1665	GlnCTG	−1.57
8	trna1243	HisGTG	−1.61
8	trna648	GlnCTG	−1.66
8	trna453	LysCTT	−1.76
8	trna552	AlaAGC	−1.83

(B) | Ten most differentially expressed miRNAs by dpi.

DPI	GeneID	log_2(FC)
1	ssc-miR-664-5p	1.21
1	ssc-miR-1306-5p	1.10
1	ssc-miR-145-5p	1.04
1	ssc-miR-296-5p	1.03
1	ssc-miR-18a	0.95
1	ssc-miR-192	−1.05
1	ssc-miR-144	−1.09
1	ssc-miR-148b-3p	−1.28
1	ssc-miR-142-3p	−1.38
1	ssc-miR-374b-5p	−1.57
3	ssc-miR-142-3p	1.17
3	ssc-miR-7	1.08
3	ssc-miR-1839-5p	0.92
3	ssc-miR-21	0.80
3	ssc-miR-142-5p	0.79
3	ssc-miR-125b	−0.76
3	ssc-miR-139-5p	−1.01
3	ssc-miR-99a	−1.18
3	ssc-miR-27b-3p	−1.91
3	ssc-miR-7134-5p	−2.23
8	ssc-miR-145-5p	1.50
8	ssc-miR-10b	1.28
8	ssc-miR-27b-3p	1.28
8	ssc-miR-9841-3p	1.28
8	ssc-miR-125b	1.23
8	ssc-miR-744	−1.17
8	ssc-let-7e	−1.20
8	ssc-miR-1285	−1.33
8	ssc-miR-129a-5p	−1.35
8	ssc-miR-296-3p	−1.59

All miRNA and tRNA log_2FC values were statistically significant based on a FDR of $q \leq 0.15$.

Pathway and Gene Ontology Analysis

Downstream analysis of miRNA gene and biological pathway targets was carried out using miRBase (Griffiths-Jones, 2006; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011; Kozomara and Griffiths-Jones, 2014) and the DIANA-TOOLS web portal (Vlachos et al., 2015b). All porcine miRNAs used were mature sequences and were first divided into up- or downregulated groups and converted to their human homolog prior to pathway analysis using the MirPath v.3 tool (Vlachos et al., 2015b). Only human sncRNA homologs and pathways were used in the comparisons. All subsequent pathways and G.O. relate to human molecular functions and biological processes for the networks or genes being targeted. The number of genes targeted for each miRNA list was based on information from the TarBase miRNA database (Vlachos et al., 2015a; Paraskevopoulou et al., 2016). The pathway and G.O. analysis used a statistical significance threshold of $q \leq 0.05$ based on the Benjamini and Hochberg FDR adjustment for multiple gene set corrections.

Noncoding RNA Interaction Analysis

The noncoding RNA (ncRNA)–protein interaction analysis was based on conversion of tRNAs to their human homologs. Swine tRNAs were matched using BLASTN to find matching human sequences and gene names prior to using the RAIN module of the RTH database site to predict possible interactions of other ncRNAs and/or mRNAs (Junge et al., 2017). Predicted interactions were presented within STRING DB (Szklarczyk et al., 2015) using the evidence view and have a confidence score of 0.2 or higher with a maximum of 10 predicted interactors allowed.

RESULTS

Clinical Evaluation of Infection

Clinical evaluation of the infected animal was examined through an analysis of viral titers. The analysis showed that within the treated samples, viral titers were present at 1 dpi and increased at every time point indicating the success of the HP-PRRSV to replicate. The viral titers for the control samples were considered to be zero, as no replication was detectable (**Figure 1**).

miRNA and tRNA Differential Expression

The outcome of our study provided differentially expressed (DE) miRNA and tRNA totals for each dpi that were statistically significant at fdr of $q \leq 0.15$ except the 1-dpi tRNA results. For the miRNAs, we observed 41 in total at 1 dpi, 14 in total at 3 dpi, and 33 in total at 8 dpi. For the tRNAs, we observed no statistically significant differential expression for 1 dpi and a total of 20 and 40 for 3 and 8 dpi, respectively.

Table 1A shows the list of statistically significant tRNAs from the interaction of treatment and time. The tRNA differential expression

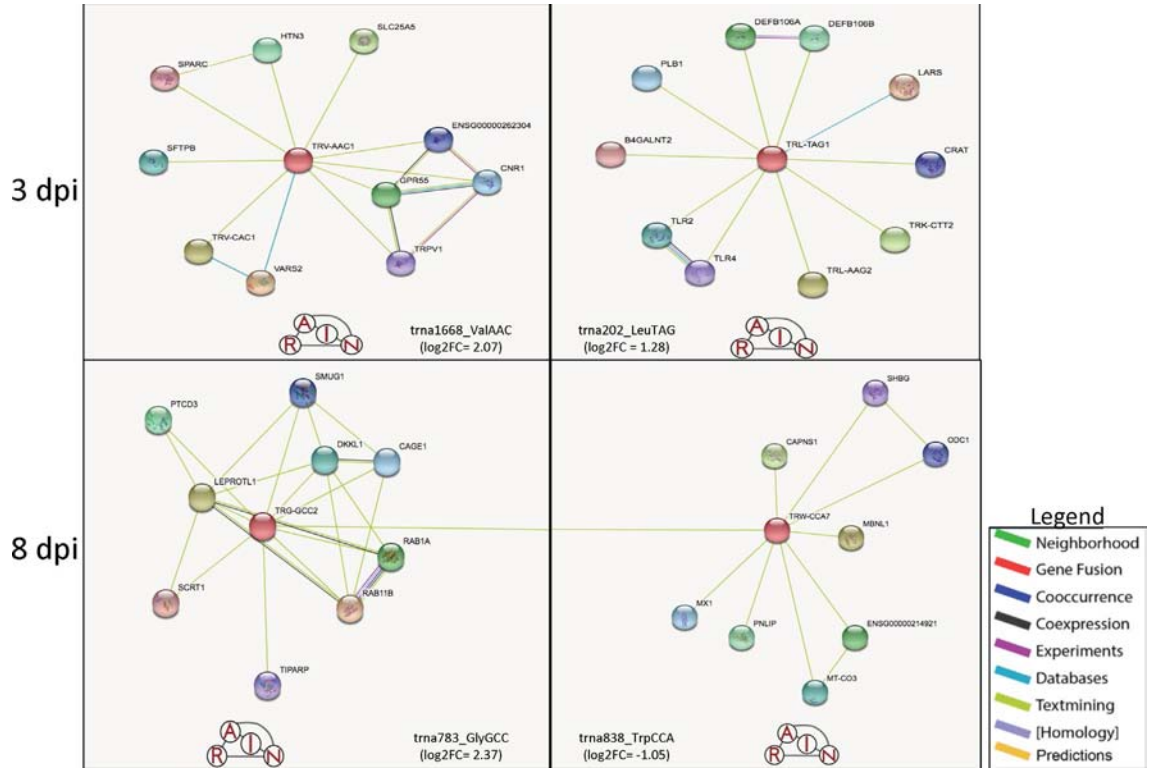


FIGURE 2 | Predicted tRNA:gene interactions for 3 and 8 dpi. The predicted interactions at 3 dpi shows that tRNA_ValAAC could be linked to signaling for inflammation and pain through its connection to *SFTPB* and *TRPV1*, while tRNA_LeuTAG is possibly linked to viral recognition through TLR2 and TLR4 interactions (top). At 8 dpi, tRNA_GlyGCC is predicted to interact with tRNA_TrpCCA, which together are predicted to interact with multiple anti-viral and apoptotic genes such as *TIPARP, MX1,* and *CAPNS1*. Figure created using the RAIN software and adapted for inclusion.

increased in both magnitude and count as dpi increased. The most abundant tRNA amino acid was alanine (n = 8) at 3 dpi with seven of the eight DE alanine tRNA genes being downregulated (**Supplementary Table 1**). The trna1668_ValAAC (log_2FC = 2.07) was the most upregulated, and trna1503_ProAGG (log_2FC = −1.78) was the most downregulated. From 3 dpi, trna1668_ValAAC is predicted to form interaction networks with Surfactant protein B (*SFTPB*), which has a homeostatic effect, as it can support alveolar functions by fostering stability within peripheral air spaces (Whitsett et al., 2010) and the gene Transient receptor potential cation channel, subfamily V, member 1 (*TRPV1*) that is involved in the intercession of inflammatory stressors such as pain (Wang et al., 2018). Also at 3 dpi was the trna202_LeuTAG (log2FC = 1.28), which forms predicted interactions with Toll-like receptors 2 and 4 (*TLR2* and *TLR4*), involved in the monocytic cell ability to recognize antigens and signal pro-inflammatory cytokines and the gene beta-1,4-N-acetyl-galactosaminyl transferase 2 (*B4GALNT2*), also known as *B4GALT*, which is a member of the Beta 4-glycosyltransferase gene family related to the gene beta-1, 4 Galactosyltransferase V (*B4GALT5*) that is shown to also be a target of the multiple miRNAs differentially expressed during this study (Jenuth, 2000; Junge et al., 2017; The UniProt, 2017) (**Figure 1**).

For 8 dpi, glycine (n = 10) was the most abundant tRNA amino acid (**Supplementary Table 1),** with trna783_GlyGCC (log_2FC = 2.37) being the highest upregulated tRNA and trna552_AlaAGC (log_2FC = −1.83) being the most downregulated (**Table 1A**). The tRNA_GlyGCC is predicted to form interactions with the gene Leptin receptor overlapping transcript-like 1 (*LEPROTL1*), a growth hormone receptor highly expressed in porcine lung tissue (Jenuth, 2000; Demarchi et al., 2007); TCDD-inducible poly(ADP-ribose) polymerase (*TIPARP*), a host defensive gene able to detect mitochondrial damage and bind viral RNA (Kozaki et al., 2017); and RAB1A member RAS oncogene family (*RAB1A*), which is involved in the biological processes of autophagy, IL-8 secretion, and post-translational protein modification (Jenuth, 2000). The tRNA_GlyGCC is also predicted to interact with trna838_TrpCCA (log_2FC = −1.05) (**Supplementary Table 1**), which is itself predicted to form interactions with the antiviral gene Myxovirus (influenza virus) resistance 1 (*MX1*) and Calpain small subunit 1 (*CAPNS1*), which is involved in the biological processes of autophagy, apoptosis, cell adhesion, and extracellular matrix degradation (Jenuth, 2000; Demarchi et al., 2007; The UniProt, 2017) (**Figure 2**). **Table 1B** shows a list of the top 10 most differentially expressed miRNAs from each dpi, some of which have been implicated as host immunomodulators during PRRSV infections as well as other viral and nonviral (Podolska et al., 2012; Maes et al., 2016; O'Leary et al., 2016; Rosenberger et al., 2017) affronts to the homeostatic state of the porcine host. The most

TABLE 2 | Unique KEGG pathways targeted differentially expressed miRNAs.

DPI	KEGG pathways targeted by upregulated miRNAs	Key targets within pathways	Classification
1	*Focal adhesion* (hsa04510)	*CAV2, AKT1, COL4A1, SHC3*	Structural
1	*Proteoglycans in cancer* (hsa05205)-4 total pathways	***Hyaluronan(HA) =*** *AKT1, SLC9A1;* ***Chondroitin sulfate/ dermatan sulfate(CSPG/DSPG) =*** *THBS1, MMP9, CD63, AKT1, EGFR;* ***Keratan sulfate(KSPG) =*** *SMAD2, MDM2, TP53, LUM;* ***Heparan sulfate(HSPG) =*** *MMP9, ITGAV, AKT1, EGFR, ITGA5, VEGFA*	Structural
1	*FoxO signaling pathway* (hsa04068)	*EGFR, PTEN, AKT1, BCL6, ATM*	Immune/Metabolism
1	*p53 signaling pathway* (hsa04115)	*ATM, TP53, PTEN,*	Immune
1	*Adherens junction* (hsa04520)	*EGFR, MET, CTNNB1*	Structural
1	*Notch signaling pathway* (hsa04330)	*ADAM17,*	Homeostatic
3	*PI3K-Akt signaling pathway* (hsa04151)	*AKT3, OSMR, PTEN, PIK3CD*	Immune/Metabolism
3	*B cell receptor signaling pathway* (hsa04662)	*BCL10, NFATC3, IKBKB, NFKBIA, NFKB1*	Immune
3	*ECM–receptor interaction* (hsa04512)	*THBS1, TNC, ITGAV, CD44, CD47*	Structural/ Immune
3	*NF-kappa B signaling pathway* (hsa04064)	*IRAK1, PTGS2, BCL10, XIAP, NFKB1, NFKB2, NFKBIA, IKBKB, TAB2*	Immune
8	*Fc gamma R-mediated phagocytosis* (hsa04666)	*AKT1, INPPL1, DNM2, CFL2, MAPK1*	Immune
8	*MAPK signaling pathway* (hsa04010)	*PDGFRB, AKT1, NFATC3, TP53, TAB2, MAP3K7, MAPK1*	Immune
8	*ECM–receptor interaction* (hsa04512)	*TNC, THBS2, ITGB1, ITGAV, FN1, LAMC1*	Structural/ Immune
DPI	**KEGG pathways targeted by downregulated miRNAs**	**Key targets within pathways**	**Classification**
1	*ECM–receptor interaction* (hsa04512)	*THBS2, CD44, CD47, ITGAV*	Structural/ Immune
1	*B cell receptor signaling pathway* (hsa04662)	*CD81, VAV3, IFITM1, NFATC3, BCL10, MALT1, CHUK*	Immune
1	*AMPK signaling pathway* (hsa04152)	*PRKAG2, PFKFB3, MAP3K7, TSC1, MTOR, ULK1*	Metabolism
1	*NF-kappa B signaling pathway* (hsa04064)	*TLR4, IL1B, VCAM1, PTGS2, BIRC2, NFKB1, ICAM, BCL2, TNFAIP3*	Immune
1	*Fc gamma R-mediated phagocytosis* (hsa04666)	*VAV3, DOCK2, PTPRC, INPPL1, DNM2*	Immune
1	*VEGF signaling pathway* (hsa04370)	*VEGFA, PTGS2, PTK2,*	Signaling
1	*TNF signaling pathway* (hsa04668)	*SOCS3, IL1B, IL6, TNFAIP3, BIRC2, VCAM1, PTGS2, MMP9, CSF1, CXCL10, ATF4, LIF*	Immune
1	*Apoptosis* (hsa04210)	*TNFSF10, CAPN2, CASP3, IL1B, IL1R1*	Homeostatic/ Metabolism
1	*Influenza A* (hsa05164)	*XPO1, RSAD2, TLR4, SOCS3, TBK1, IRF3*	Immune
3	*Focal adhesion* (hsa04510)	*VEGFA, PTK2, PDPK1*	Structural
3	*Fc gamma R-mediated phagocytosis* (hsa04666)	*ARF6, MARCKS, MAPK1, WASF2*	Immune
3	*AMPK signaling pathway* (hsa04152)	*ADIPOR2, PRKAA1, FOXO3, HMGCR, PDPK1, TSC1*	Metabolism
3	*Glycosaminoglycan biosynthesis - keratan sulfate* (hsa00533)	*B4GALT1*	Structural
3	*Mucin type O-Glycan biosynthesis* (hsa00512)	*B4GALT5, GCNT1, C1GALT1C1*	Metabolism
3	*HIF-1 signaling pathway* (hsa04066)	*HIF1A, CAMK2D*	Homeostatic
8	*Glycosaminoglycan biosynthesis—keratan sulfate* (hsa00533)	*B4GALT1*	Structural
8	*VEGF signaling pathway* (hsa04370)	*VEGFA, KDR, AKT2, PTGS2*	Signaling
8	*Wnt signaling pathway* (hsa04310)	*WNT9A, PLCB4, CTNNB1,*	Immune
8	*Apoptosis* (hsa04210)	*BCL2, AKT2, TNFSF10, IL1A, MYD88, FADD, IRAK1*	Homeostatic/ Metabolism
8	*PI3K-Akt signaling pathway* (hsa04151)	*PTEN, AKT2, F2R, IL6R, IRS1, FOXO3*	Immune/Metabolism
8	*Glycosaminoglycan biosynthesis - heparan sulfate / heparin* (hsa00534)	*EXTL3, NDST2*	Structural
8	*HIF-1 signaling pathway* (hsa04066)	IL6R, STAT3, VHL, CUL2, LTBR, IL6, TLR4	Homeostatic

The upregulated miRNAs targeted genes within the pathways for inhibition. This inhibition affected mostly structural pathways at 1 dpi, immune functions at 3 dpi, and a combination of both at 8 dpi. The downregulated miRNAs targeted genes for activation within the pathways. The activated pathways were mostly involved in immune functions at 1 and 8 dpi and structural integrity at 3 dpi. Key pathway targets based on miRNA targets within listed KEGG pathways. Analysis of statistical significance threshold set at an FDR of $q \leq 0.05$ for all pathways listed.

common differentially expressed miRNAs shared by all dpi were ssc-miR-125b (upregulated at 1 and 8 dpi and downregulated at 3 dpi) and ssc-miR-361-3p (all upregulated) (**Supplementary Table 1**).

In relation to previous PRRSV studies of sncRNA, we saw a differential expression across multiple dpi for miRNAs such as ssc-miR-142-3p (1 and 3 dpi), predicted to target the gene guanylate binding protein 5 (*GBP5*), which is shown to harbor a variant that confers some resistance to lowly pathogenic PRRSV strains (Koltes et al., 2015), and also ssc-miR-125b (3 and 8 dpi), an miRNA with anti-viral properties specific to PRRSV (Wang et al., 2013). Our study also uncovered miRNA kinetics that occurred during the 8-day experiment that were novel to the interaction of HP-PRRSV and its host *in vivo*. In upregulated miRNAs such as ssc-miR-664-5p (1 dpi), which has previously shown to become upregulated in bacterially infected lung samples (Podolska et al., 2012), ssc-miR-145-5p (8 dpi) has been revealed to be a potent inhibitor of inflammatory cytokines through its targeting of *CD40*, can prime macrophages in a M2-like manner, and is implicated

in reducing lung inflammation in humans suffering from COPD (Guo et al., 2016; O'Leary et al., 2016; Shinohara et al., 2017; Yuan et al., 2017). There was also ssc-miR-144 (1 dpi), which has been indicated as an inhibitor of the anti-viral response to influenza, another major viral respiratory disease that affects swine. There is also evidence that mature form of miR-144 also functions to suppress autophagy within host macrophages (Guo et al., 2017; Rosenberger et al., 2017). Likewise, discerned during this study were 2 miRNA groupings, the miR-30 and the miR-142 families of miRNAs that stood out across the dpi that appeared to be of importance. The miR-30 family of miRNAs appeared at least once on every dpi and was downregulated at all time points except 3 dpi, while the miR-142 family appeared downregulated at 1 and 2 upregulated at 3 and 1 at 8 dpi (**Table 1**) (**Supplementary Table 1**).

Overall Common KEGG Pathways Analysis

The pathway and G.O. analysis were examined for results that were unique to either upregulated miRNAs or downregulated miRNAs (**Table 2**) or shared between the two (data not shown). The results for the 1-, 3-, and 8-dpi infected samples showed statistical significance ($q \leq 0.05$) for multiple G.O. terms and KEGG pathways involved in the maintenance of the extracellular matrix and receptor interactions as well as various immune function-related pathways. Overall, the most prevalent shared pathway was the TGF-beta signaling pathway (*hsa04350*), a pro-inflammatory pathway, which was targeted on all dpi by the up- and downregulated miRNAs. The second most common pathway was the proteoglycans in cancer pathway (*hsa05205*). The proteoglycans in cancer pathway consisted of four total pathways: hyaluronan (*HA*), chondroitin sulfate/dermatan sulfate (*CSPG/DSPG*), keratan sulfate (*KSPG*), and heparan sulfate (*HSPG*). This composite of proteoglycan pathways is involved in structural integrity, extracellular matrix, and receptor-ligand binding and appeared as one of the most statistically significant pathways for each dpi as either a common (3 and 8 dpi) or unique (1 dpi) pathway. Other pathways shared across multiple dpi by both up- and downregulated miRNAs included endocytosis (*hsa04144*), bacterial invasion of epithelial cells (*hsa05100*), adherens junction (*hsa04520*), and mTOR signaling pathway (*hsa04150*).

miRNA Pathway Target Prediction and G.O. Analysis

The pathways that were targeted by the upregulated vs. downregulated miRNAs (**Table 2**) allowed us to observe which functional and biological processes were possibly being altered by the inhibition or activation of genes involved in these processes within the host. At 1 dpi, the unique pathways targeted by the upregulated miRNAs were mostly key structural and signaling KEGG pathways that would appear to assist viral entry. This included pathways such as focal adhesion (*hsa04510*), proteoglycans in cancer pathway (*hsa05205*), and the notch signaling pathway (*hsa04330*). In regards to the downregulated miRNAs expressed at 1 dpi, the unique pathways included key immune-related KEGG pathways related to inflammatory immune functions during PRRSV infection. This included traditional innate immune response pathways such as the TNF signaling pathway (*hsa04668*), NF-kappa B signaling pathway (*hsa04064*), the AMPK signaling pathway (*hsa04152*) capable of sensing metabolic stressors to host homeostasis, and the multifunction ECM–receptor interaction (*hsa04512*) pathway (Jenuth, 2000; Kanehisa et al., 2016; Kanehisa et al., 2017; The UniProt, 2017).

At 3 dpi, a switching of expression profiles was observed. For the upregulated miRNAs, there was an increase in the number of immune-related KEGG pathways that were targets of inhibition observed at this time point. Pathways included in this group were: NF-kappa B signaling pathway (*hsa04064*) and ECM–receptor interaction (*hsa04512*) that were activated on the previous dpi, and the PI3K-Akt signaling pathway (*hsa04151*) involved in metabolism and apoptosis. The theme of the pathways unique to the downregulated miRNAs also appeared to switch at 3 dpi. For the downregulated miRNAs, emphasis switched from mostly immune regulated to a higher quantity of structural-related pathways with roles in viral entry of respiratory tissue types such as focal adhesion (*hsa04510*), Mucin type O-Glycan biosynthesis (*hsa00512*), and Glycosaminoglycan biosynthesis—keratan sulfate (*hsa00533*) (Jenuth, 2000; Souza-Fernandes et al., 2006; Kanehisa et al., 2016; Kanehisa et al., 2017; The UniProt, 2017). By 8 dpi, the pathways being targeted by the up- and downregulated miRNAs continued to display very kinetic profiles. The upregulated miRNAs now targeted pathways that involved both key structural- and immune-related functions, while the downregulated miRNAs targeted pathways continued to place emphasis again on host immune functions like apoptosis (*hsa04210*) and the Wnt signaling pathway (*hsa04310*) (Pecina-Slaus, 2010; Kanehisa et al., 2016; Villasenor et al., 2017) (**Table 2**).

Additionally, a gene ontology (G.O.) analysis was conducted for the upregulated miRNAs in order to facilitate a better understanding of what biological and molecular functions were being inhibited by the overexpression of the different miRNAs. A Venn diagram (not shown) was used to filter the G.O. terms and determine which were common or unique to each time point in an attempt to elucidate which biological functions of the host were the most attenuated by miRNA overexpression. The terms that were shared across all time points (n = 127) included a majority of host immune processes generally related to viral–host interactions during infections. The shared list contained the G.O. terms negative regulation of type I interferon production (GO:0032480), extracellular matrix disassembly (GO:0022617), and virus receptor activity (GO:0001618).

At 1 dpi, the unique G.O. terms (n = 20) included antigen processing and presentation of peptide antigen *via* MHC class I (GO:0002474), ncRNA metabolic process (GO:0034660), negative regulation of transforming growth factor beta receptor signaling pathway (GO:0030512), and tRNA metabolic process. At 3 dpi, the counts of unique terms dropped considerably (n = 6) from 1 dpi and highlighted terms related to sensing of biotic and abiotic stressors and inhibition by small RNA functions with the terms cytoplasmic stress granule (GO:0010494) and negative regulation of translation (GO:0017148). The unique G.O. terms associated with the miRNA upregulation at 8 dpi (n = 41) were the largest grouping of terms of all time points and encompassed multiple host immune and metabolic function-related terms. Some of

the key terms from 8 dpi included collagen catabolic process (GO:0030574), O-glycan processing (GO:0016266), and type I interferon signaling pathway (GO:0060337). The G.O. terms that were unique to the downregulated miRNAs at each dpi were also examined and showed groupings of terms that indicated targeting of various immune functions (not shown).

DISCUSSION

Interaction of Treatment and Dpi Reveals Pro-Viral and Anti-Viral Battle Over Host Pathways

Examination of the results indicated that the miRNAs and tRNAs displayed expression patterns that seem to suggest they behave in a manner that both promotes and fights HP-PRRSV infection. This battle for post-transcriptional control over the regulation of host gene expression during infection appeared to take place on several fronts. These fronts were highlighted by the differential expression of candidate miRNAs, tRNAs, and pathways related to several different host biological groupings. The groupings are best represented by the relationship of the pathways they contain, such as structural-related networks that consist of pathways that affect structural integrity and receptor binding. Additionally, there are the immune function-related pathways that spotlight the tug-of-war between host and pathogen. Lastly, there are the pathways that control metabolic activity, which collectively point out the perturbation the virus causes to host homeostasis. Even more striking is that these biological groupings are unique to either the overexpressed or under-expressed miRNAs and reinforced by the ncRNA:mRNA interactions predicted for some of the tRNA genes (Banerjee et al., 2013; Bai et al., 2015).

HP-PRRSV Infection Stimulated Differential Expression of Key Extracellular miRNAs

Analysis of the miRNA results yielded a group of candidates from each dpi that appeared to be linked to the immunosuppressive effects of the HP-PRRSV infection. Many of these were upregulated miRNAs listed as residing in the extracellular space of cells and may possibly highlight the usurping of host extracellular miRNAs by HP-PRRSV to facilitate viral entry/replication. The upregulation was matched by downregulation of additional miRNAs, likely still under host control, which could help boost the activity and expression of immune-related pathways. The direction of expression for the miRNAs were closely tied to the dpi, as many were observed to be overexpressed at one time point, only to be under-expressed at another. Despite this, the results yielded candidate miRNA's with potential anti- and pro-viral actions during HP-PRRSV infections of swine. From the upregulated miRNAs at 1, 3, and, 8 dpi were miRNAs Ssc-miR-145-5p, Ssc-miR-142-3p, and Ssc-miR-125b, respectively.

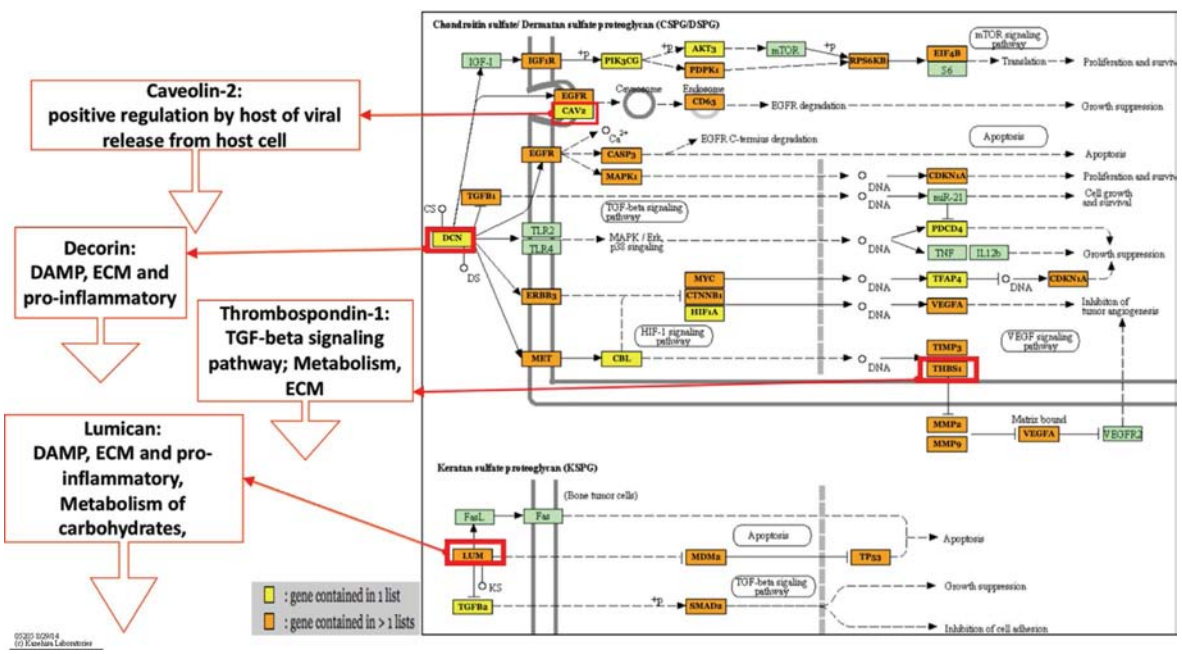


FIGURE 3 | Proteoglycans in cancer-predicted pathway. This figure highlights two of the four proteoglycans in cancer pathways, chondroitin sulfate/dermatan sulfate (*CSPG/DSPG*) and keratan sulfate (*KSPG*). These networks show that there are multiple genes (orange and yellow boxes) related to cytokine signaling and viral entry that are being targeted for inhibition that would cause host dysregulation and impair the ability to properly respond or maintain homeostasis during infection. Additionally, genes shown in previous HP-PRRSV transcriptome studies to be downregulated are shown to be targets of the upregulated miRNAs. *DCN* and *LUM* are DAMPs, which lead to downregulation of inflammatory DAMP signals and downstream *TGF-Beta* signaling, which is involved in anti-viral immunity. **Figure 3** shows upregulated miRNAs only from all time points. Figure created using Mirpath V3 software and adapted for inclusion. Figure based on KEGG pathways. Figure legend refers to if gene in pathway has 1 (yellow) or >1 (orange) miRNAs targeting it.

The miRNA Ssc-miR-125b (upregulated at 1 and 8 dpi only) along with Ssc-miR-361-3p (upregulated at all dpi) were the only miRNAs to appear statistically significant at each dpi. The miRNA Ssc-miR-145-5p has been shown to be an inhibitor of inflammatory cytokines signaling with the ability to cause alternative macrophage priming that accents anti-inflammatory signaling within the host (Guo et al., 2016; Yuan et al., 2017). This may be evidence of an early ability of PRRSV to suppress innate immune functions or signaling by overriding the M1 priming that would initiate host pro-inflammatory signaling. Also, of interest at 1 dpi was the downregulated Ssc-miR-144, which in its mature form can suppress proper macrophage functions and is also considered an inhibitor of the anti-viral response to influenza, another major porcine respiratory disease (Rosenberger et al., 2017). The difference in the previously studied functions of Ssc-miR-145-5p and Ssc-miR-144 show that at 1-dpi expression of the miRNAs is gauged in both pro- and anti-viral directions.

Upregulated at 3 dpi was the miRNA Ssc-miR-142-3p, which specifically when upregulated has been shown to impair the ability of monocyte-derived cells to properly present and process antigens to the adaptive arm of the immune system (Naqvi et al., 2016). This gives insight into a possible mechanism that HP-PRRSV uses for stalling the host adaptive immune response. Also observed in the study were other differentially expressed members of the miR-142 family of miRNAs. Also upregulated at 3 dpi was the related Ssc-miR-142-5p, which is predicted to target porcine guanylate-binding protein 5 (*GBP5*) gene, which contains the WUR SNP, a low pathogenic PRRSV resistance variant (Koltes et al., 2015). The 3-dpi overexpression of these miRNAs may be evidence of possible mechanisms available to HP-PRRSV strains to stall host immune response and possibly cancel out any protection offered by the WUR variant through GBP5 silencing.

The upregulation of miRNAs at 1 and 3 dpi seemed to favor pro-viral activity within the host; however, by 8 dpi, miRNA upregulation was more anti-viral. This was observed in the host overexpression of Ssc-miR-125b, an extracellular miRNA that is computationally predicted to limit PRRS viral levels through suppressive targeting of host NF-kB signaling (Qureshi et al., 2014). These phenomena may be related to the active role the miRNA plays in the regulation of monocytic cell inflammatory signaling (Duroux-Richard et al., 2016).

tRNA Differential Expression During HP-PRRSV Infection Indicates Change in Host Homeostasis Benefitting HP-PRRSV

The tRNA differential expression was not significant at 1 dpi but was statistically significant at 3 and 8 dpi after HP-PRRSV infection. Interestingly, the pattern of statistically significant tRNAs and their magnitudes of expression (**Supplementary Table 1**) appear to follow the trend seen in the viral load that increased with dpi (**Figure 1**). The changes in tRNA expression overtime are possibly the result of the virus modulating tRNA expression through hijacking host cellular resources. This change in host resources can then lead to changes in the metabolism of the host that eventually disrupts the proper activation of monocytic cells such as macrophages, allowing the virus to proceed unimpeded by the host cytokines (Langston et al., 2017) toward entry and proliferation. In a 2016 paper, Rappe et al. (2016) was able to observe that PRRS type 1 and 2 virus nucleocapsids can contain an amino acid substitution that replaces a threonine with alanine causing the PRRS virus strain in that study to escape immune system protection. Therefore, it is possible that the virus is promoting differential expression of host alanine as a means of both hijacking needed nutrients and avoiding detection by innate immunity. There is also a possibility that the consistent downregulation of tRNA alanine genes at 3 and 8 dpi (**Supplementary Table 1**) is host initiated to hasten cellular degradation to limit host resources to the virus.

The trna1668_ValAAC was predicted to form interaction networks with *SFTPB* and *TRPV1* (**Figure 2**), two genes involved in processing host signals for inflammation and pain that could also be linked to the changes in miRNA expression at 3 dpi that are shifting the host immune response. Additionally, at 3 dpi is the predicted interaction of the trna202_LeuTAG with the Beta 4-glycosyltransferase gene family that also contains genes involved in the functioning of chondroitin, a major proteoglycan pathway shown to be highly targeted by the miRNAs at each dpi. This can be seen in the results at 3 dpi where miRNAs ssc-miR-27b-3p and ssc-miR-23a-3p targeting *B4GALT5* (Zhang et al., 2018) are being downregulated in the mucin type-o pathway in a possible attempt to bolster mucosal immunity in the host airway.

The prevalence of tRNA glycine genes at 8 dpi could be linked to the role glycine plays in the biochemical composition of collagen (Li and Wu, 2018), a tissue type composed of genes such as *COL4A1* and *COL5A2* shown in previous studies to be dysregulated during HP-PRRSV infections (Miller et al., 2017), and targeted by upregulated miRNAs in different pathways during our study (**Table 2**). Glycine is one of the key components in collagen, and collagen is a key component of the ECM, which our results showed to be compromised by multiple upregulated miRNAs (**Table 1**) (**Supplementary Table 1**). Additionally, glycine and its metabolites are needed by mammals like swine to support proper immune functions. However, the upregulation at 8 dpi could be more closely related to glycine's ability as an anti-inflammatory molecule.

KEGG Pathways Analysis Indicates Coupled Viral Entry and Immunosuppressive Effect of HP-PRRSV

The pathways that were shared across all dpi such as the TGF-beta signaling pathway is a continuation of the suppression observed with the four proteoglycans in cancer pathways. The DE miRNAs within the TGF-beta signaling pathway likely perturbs multiple transcription and cofactors leading to immunosuppression, delayed apoptotic induction, and ECM dysregulation that is also shown to be affected. The first front of the battle between host and virus appears to take place at 1 dpi within inhibited structural pathways, concomitant with viral entry and proliferation, and activated immune response pathways that support pro-inflammatory signaling (**Table 2**). The other common pathway, the proteoglycans

in cancer pathway, was actually unique to 1 dpi (**Figure 3**) and revealed a class of miRNA-targeted structural genes that also functions as part of the innate immunity collectively referred to as damage-associated molecular pattern signals or DAMPs. Two DAMPs in particular, decorin (*DCN*) and lumican (*LUM*), have been shown to be heavily differentiated in HP-PRRSV infections (Miller et al., 2017) in a previous study and were targets of multiple upregulated miRNAs within our results. It is possible that miRNA overexpression reduces the ability of proteoglycan DAMPs like *DCN* and *LUM* to promote inflammatory cytokine signaling (Merline et al., 2011; Moreth et al., 2012), while also weakening the ECM to help promote viral invasion and proliferation.

The pathways targeted for miRNA-directed suppression at 1 dpi such as focal adhesion and the proteoglycans in cancer pathway could indicate early viral manipulation of host resources linked to documented viral entry strategies involving the binding of host glycoprotein molecules, which have many identities such as cell-surface and transmembrane receptors that make up extracellular matrix proteins such as proteoglycans and integrins (Rabinovich et al., 2012; Cossart and Helenius, 2014). These pathways are crucial to the makeup of the extracellular matrix and may indicate why the ECM–receptor interaction (*hsa04512*) pathway is specifically targeted for activation by the downregulated miRNAs. In line with the viruses' infection and proliferation at 1 dpi, the host was not only bolstering defense to viral entry with increased ECM–receptor activity but also innate immune responses for cytokine signaling possible through activity of the TNF signaling pathway (*hsa04668*) and the NF-kappa B signaling pathway (*hsa04064*). There is also some indication that the host metabolism is starting to become perturbed at 1 dpi due to activation of genes within the AMPK signaling pathway (*hsa04152*) that is involved in metabolic functions, especially in low energy states (Mihaylova and Shaw, 2011).

By 3 dpi, the unique up- and downregulated miRNAs had switched targeting roles with respect to the biological processes that appeared to be either inhibited or activated within the host. The upregulated miRNA was now inhibiting immune function pathways such as the ECM–receptor interaction pathway (*hsa04512*) and NF-kappa B signaling pathway (*hsa04064*) that were previously being activated. The switch from structural to immune pathway inhibition may serve as possible indicators of the dysregulation of monocytic cell immune functions experienced by infected macrophages to protect and prolong viral replication. Also supporting the idea that HP-PRRSV had switched to modulating host immune pathways to support survival was the observance of the PI3K-Akt signaling pathway (*hsa04151*) inhibition. This pathway is known to be involved in cell survival and can be usurped by viral pathogens to promote survival (Dunn and Connor, 2012); therefore, the host may be regulating miRNA expression against the pathway in an attempt to reduce usurpation by HP-PRRSV. Another suggestion is that the inhibition of the PI3K-Akt signaling pathway could be linked to the 1-dpi inhibition of the FoxO signaling pathway (*hsa04068*) and is a means for HP-PRRSV to lower the apoptotic functions for these pathways to promote self-survival against host immune functions. The unique pathways based on the downregulated miRNAs at 3 dpi had now begun to target more structural and metabolic pathways such as the Mucin type O-Glycan biosynthesis (*hsa00512*), AMPK signaling pathway (*hsa04152*), and the HIF-1 signaling pathway (*hsa04066*). It is uncertain whether this potential increase in structural and metabolic activity is anti- or pro-viral in nature; however, it does show evidence that host homeostasis is being dysregulated. If it is anti-viral in nature, the pathway activity may be supporting repair of components such as the ECM after viral entry. However, if it is pro-viral, it might hint at AMPK and HIF-1 signaling pathway involvement in metabolic processes that can detract from normal macrophage activation and may be connected to the 8 dpi upregulation of glycine tRNAs based on *HIF1A*'s role in glycolysis (Langston et al., 2017).

By 8 dpi, the affected unique pathways had flipped roles once again in regard to the biological clustering of inhibited and activated pathways. The 8-dpi inhibited pathways now incorporated a combination of structural and immunological pathways that included the ECM–receptor interaction pathway (*hsa04512*), Fc gamma R-mediated phagocytosis (*hsa04666*), and the MAPK signaling pathway (*hsa04010*). The 8-dpi pathway analysis also showed that the unique activated pathways also clustered around both immune and structural pathway indicating some importance of proteoglycans and pro-inflammatory cytokine signaling. It is also possible that the pathways targeted by the miRNA differential expression at 8 dpi is a reflection of the losing battle between host immunity and HP-PRRSV virulence related to impairment of normal macrophage functions due to PRRSV cell tropism.

CONCLUSIONS

Taken together, the pathway analyses suggest that the changes in host homeostasis were affected through the ability of HP-PRRSV to disturb host structural, immunologic, and metabolic pathways. These targeted pathways, along with the predicted tRNA:gene interactions, highlighted both inhibition and activation of pathways involved in viral entry, proliferation, and pro-inflammatory signaling that may underlie the ability of PRRSV to hinder homeostasis through sncRNA dysregulation. Small noncoding RNA (sncRNA) expression during HP-PRRSV infection can affect the differential expression of miRNAs and tRNAs and can exist as an evasion to canonical host immune responses when expressed in patterns exhibited across the experiment's time points. Highly pathogenic PRRSV appeared to have the ability to induce differential expression of both miRNAs and tRNAs as part of its pathogenic course that perturbed structural, metabolic, and immunogenic pathways. The action of these sncRNAs created post-transcriptional changes to the overall ability of the host to maintain cellular homeostasis in the presence of the pathogen.

DATA AVAILABILITY

The datasets generated and/or analyzed for this study can be found in the GEO repository, GSE121980, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121980.

ETHICS STATEMENT

The animal use protocol was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the National Animal Disease Center-USDA-Agricultural Research Service. Written informed consent to use the animals in the study was obtained from the Wilson Farms, WI.

AUTHOR CONTRIBUTIONS

LM contributed to the study conception, data collection, research design, and manuscript writing. DF contributed to the data preparation, data analysis, research design, and manuscript writing.

FUNDING

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer. This work was mainly supported by the USDA NIFA AFRI 2013-67015-21236 and in part by the USDA NIFA AFRI 2015-67015-23216. DF was supported in part by an appointment to the Agricultural Research Service Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the US Department of Energy (DOE) and the US Department of Agriculture. ORISE is managed by Oak Ridge Associated Universities under DOE contract no. DE-AC05-06OR23100.

ACKNOWLEDGMENTS

We would like to thank Dr. Susan Brockmeier and lab for help during sample collection, Dr. Kay Faaberg for access to the virus used within the study, Dr. Joan Lunney for access to the globin depletion protocol, Randy Atchison for technical support, Sarah Anderson for sample collection, sample preparation, and her excellent technical support, and Sue Ohlendorf for secretarial assistance in preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00691/full#supplementary-material

SUPPLEMENTARY TABLE 1 | tRNA and mRNA differential expression by dpi. Excel chart of the differential expression (FDR ≤ 0.15) observed for each class of sncRNA for 1, 3, and 8 days post-infection with HP-PRRSV.

SUPPLEMENTARY TABLE 2 | Unique upregulated and downregulated G.O. terms by dpi. Excel chart of statistically significant (FDR ≤ 0.05) G.O. terms based on miRNA analysis split by direction of differential expression and dpi.

SUPPLEMENTARY TABLE 3 | Filtering and mapping statistics based on size selection of reads 18–80 bps after library creation.

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., et al. (2016). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44 (W1), W3–W10. doi: 10.1093/nar/gkw343

Bai, L., Liang, R., Yang, Y., Hou, X., Wang, Z., Zhu, S., et al. (2015). MicroRNA-21 regulates PI3K/Akt/mTOR signaling by targeting TGFbetaI during skeletal muscle development in pigs. *PLoS One* 10 (5), e0119396. doi: 10.1371/journal.pone.0119396

Banerjee, S., Cui, H., Xie, N., Tan, Z., Yang, S., Icyuz, M., et al. (2013). miR-125a-5p regulates differential activation of macrophages and inflammation. *J. Biol. Chem.* 288 (49), 35428–35436. doi: 10.1074/jbc.M112.426866

Bioinformatics & Evolutionary Genomics [Online]. (2017) Available: http://bioinformatics.psb.ugent.be/webtools/Venn/ [Accessed].

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* 10, 11–21. doi: 10.1002/0471142727.mb1910s89

Bowie, A. G., and Unterholzner, L. (2008). Viral evasion and subversion of pattern-recognition receptor signalling. *Nat. Rev. Immunol.* 8, 911. doi: 10.1038/nri2436

Bruscella, P., Bottini, S., Baudesson, C., Pawlotsky, J. M., Feray, C., and Trabucchi, M. (2017). Viruses and miRNAs: more friends than foes. *Front. Microbiol.* 8, 824. doi: 10.3389/fmicb.2017.00824

Choi, I., Bao, H., Kommadath, A., Hosseini, A., Sun, X., Meng, Y., et al. (2014). Increasing gene discovery and coverage using RNA-seq of globin RNA reduced porcine blood samples. *BMC Genomics* 15 (1), 954. doi: 10.1186/1471-2164-15-954

Cossart, P., and Helenius, A. (2014). Endocytosis of viruses and bacteria. *Cold Spring Harb. Perspect. Biol.* 6 (8), 1–28. doi: 10.1101/cshperspect.a016972

Dekkers, J., Rowland, R. R. R., Lunney, J. K., and Plastow, G. (2017). Host genetics of response to porcine reproductive and respiratory syndrome in nursery pigs. *Vet. Microbiol.* 209, 107–113. doi: 10.1016/j.vetmic.2017.03.026

Demarchi, F., Bertoli, C., Copetti, T., Eskelinen, E.-L., and Schneider, C. (2007). Calpain as a novel regulator of autophagosome formation. *Autophagy* 3 (3), 235–237. doi: 10.4161/auto.3661

Duan, X., Nauwynck, H. J., and Pensaert, M. B. (1997). Effects of origin and state of differentiation and activation of monocytes/macrophages on their susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV). *Arch. Virol.* 142 (12), 2483–2497. doi: 10.1007/s007050050256

Dunn, E. F., and Connor, J. H. (2012). HijAkt: the PI3K/Akt pathway in virus replication and pathogenesis. *Prog. Mol. Biol. Transl. Sci.* 106, 223–250. doi: 10.1016/B978-0-12-396456-4.00002-X

Duroux-Richard, I., Roubert, C., Ammari, M., Presumey, J., Grun, J. R., Haupl, T., et al. (2016). miR-125b controls monocyte adaptation to inflammation through mitochondrial metabolism and dynamics. *Blood* 128 (26), 3125–3136. doi: 10.1182/blood-2016-02-697003

Fleming, D. S., and Miller, L. C. (2018). Identification of small non-coding RNA classes expressed in swine whole blood during HP-PRRSV infection. *Virology* 517, 56–61. doi: 10.1016/j.virol.2018.01.027

Griffiths-Jones, S. (2006). miRBase: the microRNA sequence database. *Methods Mol. Biol.* 342, 129–138. doi: 10.1385/1-59745-123-1:129

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34 (Database issue), D140–D144. doi: 10.1093/nar/gkj112

Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36 (Database issue), D154–158. doi: 10.1093/nar/gkm952

Guo, L., Zhou, L., Gao, Q., Zhang, A., Wei, J., Hong, D., et al. (2017). MicroRNA-144-3p inhibits autophagy activation and enhances Bacillus Calmette-Guerin

infection by targeting ATG4a in RAW264.7 macrophage cells. *PLoS One* 12 (6), e0179772. doi: 10.1371/journal.pone.0179772

Guo, X., Yu, L., Chen, M., Wu, T., Peng, X., Guo, R., et al. (2016). miR-145 mediated the role of aspirin in resisting VSMCs proliferation and anti-inflammation through CD40. *J. Transl. Med.* 14 (1), 211. doi: 10.1186/s12967-016-0961-2

Hicks, J. A., Yoo, D., and Liu, H. C. (2013). Characterization of the microRNAome in porcine reproductive and respiratory syndrome virus infected macrophages. *PLoS One* 8 (12), e82054. doi: 10.1371/journal.pone.0082054

Holtkamp, D. J., Kliebenstein, J. B., Neumann, E. J., Zimmerman, J. J., Rotto, H., Yoder, T. K., et al. (2013). Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J. Swine Health Prod.* 21, 272–284. doi: 10.31274/ans_air-180814-28

Hotamisligil, G. S. (2006). Inflammation and metabolic disorders. *Nature* 444 (7121), 860–867. doi: 10.1038/nature05485

Jenuth, J. P. (2000). The NCBI. Publicly available tools and resources on the Web. *Methods Mol. Biol.* 132, 301–312. doi: 10.1385/1-59259-192-2:301

Junge, A., Refsgaard, J. C., Garde, C., Pan, X., Santos, A., Alkan, F., et al. (2017). RAIN: RNA-protein Association and Interaction Networks. *Database (Oxford)* 2017, 1–9. doi: 10.1093/database/baw167

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44 (D1), D457–462. doi: 10.1093/nar/gkv1070

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Koltes, J. E., Fritz-Waters, E., Eisley, C. J., Choi, I., Bao, H., Kommadath, A., et al. (2015). Identification of a putative quantitative trait nucleotide in guanylate binding protein 5 for host response to PRRS virus infection. *BMC Genomics* 16, 412. doi: 10.1186/s12864-015-1635-9

Kozaki, T., Komano, J., Kanbayashi, D., Takahama, M., Misawa, T., Satoh, T., et al. (2017). Mitochondrial damage elicits a TCDD-inducible poly(ADP-ribose) polymerase-mediated antiviral response. *Proc. Natl. Acad. Sci. U. S. A.* 114 (10), 2681–2686. doi: 10.1073/pnas.1621508114

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39 (Database issue), D152–157. doi: 10.1093/nar/gkq1027

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42 (Database issue), D68–73. doi: 10.1093/nar/gkt1181

Langston, P. K., Shibata, M., and Horng, T. (2017). Metabolism supports macrophage activation. *Front. Immunol.* 8, 61. doi: 10.3389/fimmu.2017.00061

Li, P., and Wu, G. (2018). Roles of dietary glycine, proline, and hydroxyproline in collagen synthesis and animal growth. *Amino Acids* 50 (1), 29–38. doi: 10.1007/s00726-017-2490-6

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656

Maes, T., Cobos, F. A., Schleich, F., Sorbello, V., Henket, M., De Preter, K., et al. (2016). Asthma inflammatory phenotypes show differential microRNA expression in sputum. *J. Allergy Clin. Immunol.* 137 (5), 1433–1446. doi: 10.1016/j.jaci.2016.02.018

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17 (1), 10–12. doi: 10.14806/ej.17.1.200

Merline, R., Moreth, K., Beckmann, J., Nastase, M. V., Zeng-Brouwers, J., Tralhao, J. G., et al. (2011). Signaling by the matrix proteoglycan decorin controls inflammation and cancer through PDCD4 and MicroRNA-21. *Sci. Signal* 4 (199), ra75. doi: 10.1126/scisignal.2001868

Mihaylova, M. M., and Shaw, R. J. (2011). The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat. Cell Biol.* 13 (9), 1016–1023. doi: 10.1038/ncb2329

Miller, L. C., Fleming, D., Arbogast, A., Bayles, D. O., Guo, B., Lager, K. M., et al. (2012). Analysis of the swine tracheobronchial lymph node transcriptomic response to infection with a Chinese highly pathogenic strain of porcine reproductive and respiratory syndrome virus. *BMC Vet. Res.* 8, 208. doi: 10.1186/1746-6148-8-208

Miller, L. C., Fleming, D. S., Li, X., Bayles, D. O., Blecha, F., and Sang, Y. (2017). Comparative analysis of signature genes in PRRSV-infected porcine monocyte-derived cells to different stimuli. *PLoS One* 12 (7), e0181256. doi: 10.1371/journal.pone.0181256

Moreth, K., Iozzo, R. V., and Schaefer, L. (2012). Small leucine-rich proteoglycans orchestrate receptor crosstalk during inflammation. *Cell Cycle* 11 (11), 2084–2091. doi: 10.4161/cc.20316

Naqvi, A. R., Fordham, J. B., Ganesh, B., and Nares, S. (2016). miR-24, miR-30b and miR-142-3p interfere with antigen processing and presentation by primary macrophages and dendritic cells. *Sci. Rep.* 6, 32925. doi: 10.1038/srep32925

O'Leary, L., Sevinc, K., Papazoglou, I. M., Tildy, B., Detillieux, K., Halayko, A. J., et al. (2016). Airway smooth muscle inflammation is regulated by microRNA-145 in COPD. *FEBS Lett.* 590 (9), 1324–1334. doi: 10.1002/1873-3468.12168

Paraskevopoulou, M. D., Vlachos, I. S., and Hatzigeorgiou, A. G. (2016). DIANA-TarBase and DIANA suite tools: studying experimentally supported microRNA Targets. *Curr. Protoc. Bioinform.* 55, 12 14 11–12 14 18. doi: 10.1002/cpbi.12

Pecina-Slaus, N. (2010). Wnt signal transduction pathway and apoptosis: a review. *Cancer Cell. Int.* 10, 22. doi: 10.1186/1475-2867-10-22

Podolska, A., Anthon, C., Bak, M., Tommerup, N., Skovgaard, K., Heegaard, P. M., et al. (2012). Profiling microRNAs in lung tissue from pigs infected with Actinobacillus pleuropneumoniae. *BMC Genomics* 13, 459. doi: 10.1186/1471-2164-13-459

Qureshi, A., Thakur, N., Monga, I., Thakur, A., and Kumar, M. (2014). VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. *Database (Oxford)* 2014, 1–10. doi: 10.1093/database/bau103

Rabinovich, G. A., van Kooyk, Y., and Cobb, B. A. (2012). Glycobiology of immune responses. *Ann. N. Y. Acad. Sci.* 1253, 1–15. doi: 10.1111/j.1749-6632.2012.06492.x

Rappe, J. C., Garcia-Nicolas, O., Fluckiger, F., Thur, B., Hofmann, M. A., Summerfield, A., et al. (2016). Heterogeneous antigenic properties of the porcine reproductive and respiratory syndrome virus nucleocapsid. *Vet. Res.* 47 (1), 117. doi: 10.1186/s13567-016-0399-9

Rosenberger, C. M., Podyminogin, R. L., Diercks, A. H., Treuting, P. M., Peschon, J. J., Rodriguez, D., et al. (2017). miR-144 attenuates the host response to influenza virus by targeting the TRAF6-IRF7 signaling axis. *PLoS Pathog.* 13 (4), e1006305. doi: 10.1371/journal.ppat.1006305

Shanmukhappa, K., Kim, J. K., and Kapil, S. (2007). Role of CD151, A tetraspanin, in porcine reproductive and respiratory syndrome virus infection. *Virol. J.* 4, 62. doi: 10.1186/1743-422X-4-62

Shinohara, H., Kuranaga, Y., Kumazaki, M., Sugito, N., Yoshikawa, Y., Takai, T., et al. (2017). Regulated polarization of Tumor-Associated Macrophages by miR-145 *via* colorectal cancer-derived extracellular vesicles. *J. Immunol.* 199 (4), 1505–1515. doi: 10.4049/jimmunol.1700167

Souza-Fernandes, A. B., Pelosi, P., and Rocco, P. R. (2006). Bench-to-bedside review: the role of glycosaminoglycans in respiratory disease. *Crit. Care* 10 (6), 237. doi: 10.1186/cc5069

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43 (Database issue), D447–D452. doi: 10.1093/nar/gku1003

Taxis, T. M., Bauermann, F. V., Ridpath, J. F., and Casas, E. (2017). Circulating MicroRNAs in serum from cattle challenged with bovine viral diarrhea virus. *Front. Genet.* 8, 91. doi: 10.3389/fgene.2017.00091

The UniProt, C. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–D169. doi: 10.1093/nar/gkw1099

van Geelen, A. G. M., Anderson, T. K., Lager, K. M., Das, P. B., Otis, N. J., Montiel, N. A., et al. (2018). Porcine reproductive and respiratory disease virus: Evolution and recombination yields distinct ORF5 RFLP 1-7-4 viruses with individual pathogenicity. *Virology* 513, 168–179. doi: 10.1016/j.virol.2017.10.002

Villasenor, T., Madrid-Paulino, E., Maldonado-Bravo, R., Urban-Aragon, A., Perez-Martinez, L., and Pedraza-Alva, G. (2017). Activation of the Wnt Pathway by Mycobacterium tuberculosis: a Wnt-Wnt Situation. *Front. Immunol.* 8, 50. doi: 10.3389/fimmu.2017.00050

Vlachos, I. S., Paraskevopoulou, M. D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., et al. (2015a). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.* 43 (Database issue), D153–159. doi: 10.1093/nar/gku1215

Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., et al. (2015b). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* 43 (W1), W460–W466. doi: 10.1093/nar/gkv403

Wang, D., Cao, L., Xu, Z., Fang, L., Zhong, Y., Chen, Q., et al. (2013). MiR-125b reduces porcine reproductive and respiratory syndrome virus replication by negatively regulating the NF-kappaB pathway. *PLoS One* 8 (2), e55838. doi: 10.1371/journal.pone.0055838

Wang, Y., Gao, Y., Tian, Q., Deng, Q., Wang, Y., Zhou, T., et al. (2018). TRPV1 SUMOylation regulates nociceptive signaling in models of inflammatory pain. *Nat. Commun.* 9 (1), 1529. doi: 10.1038/s41467-018-03974-7

Whitsett, J. A., Wert, S. E., and Weaver, T. E. (2010). Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu. Rev. Med.* 61 (1), 105–119. doi: 10.1146/annurev.med.60.041807.123500

Xiao, Y., An, T. Q., Tian, Z. J., Wei, T. C., Jiang, Y. F., Peng, J. M., et al. (2015). The gene expression profile of porcine alveolar macrophages infected with a highly pathogenic porcine reproductive and respiratory syndrome virus indicates overstimulation of the innate immune system by the virus. *Arch. Virol.* 160 (3), 649–662. doi: 10.1007/s00705-014-2309-7

Yuan, M., Zhang, L., You, F., Zhou, J., Ma, Y., Yang, F., et al. (2017). MiR-145-5p regulates hypoxia-induced inflammatory response and apoptosis in cardiomyocytes by targeting CD40. *Mol. Cell. Biochem.* 431 (1–2), 123–131. doi: 10.1007/s11010-017-2982-4

Zhang, L., Ren, J., Shi, P., Lu, D., Zhao, C., Su, Y., et al. (2018). The immunological regulation roles of porcine beta-1, 4 Galactosyltransferase V (B4GALT5) in PRRSV infection. *Front. Cell. Infect. Microbiol.* 8, 48. doi: 10.3389/fcimb.2018.00048

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Fleming and Miller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 August 2019
doi: 10.3389/fgene.2019.00701

Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle

Joana Lima[1], Marc D. Auffret[1], Robert D. Stewart[2], Richard J. Dewhurst[1], Carol-Anne Duthie[1], Timothy J. Snelling[3], Alan W. Walker[3], Tom C. Freeman[2†], Mick Watson[2] and Rainer Roehe[1]**

[1] Beef and Sheep Research Centre, Future Farming Systems Group, Scotland's Rural College, Edinburgh, United Kingdom, [2] Division of Genetics and Genomics, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, [3] The Rowett Institute, University of Aberdeen, Aberdeen, United Kingdom

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Robert W. Li, Agricultural Research Service (USDA), United States
Prakash G. Koringa, Anand Agricultural University, India

***Correspondence:**
Joana Lima
Joana.Lima@sruc.ac.uk
Rainer Roehe
Rainer.Roehe@sruc.ac.uk

†ORCID:
Tom Freeman
orcid.org/0000-0001-5235-8483

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 07 February 2019
Accepted: 03 July 2019
Published: 08 August 2019

Citation:
Lima J, Auffret MD, Stewart RD, Dewhurst RJ, Duthie C-A, Snelling TJ, Walker AW, Freeman TC, Watson M and Roehe R (2019) Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle. Front. Genet. 10:701. doi: 10.3389/fgene.2019.00701

The rumen microbiome is essential for the biological processes involved in the conversion of feed into nutrients that can be utilized by the host animal. In the present research, the influence of the rumen microbiome on feed conversion efficiency, growth rate, and appetite of beef cattle was investigated using metagenomic data. Our aim was to explore the associations between microbial genes and functional pathways, to shed light on the influence of bacterial enzyme expression on host phenotypes. Two groups of cattle were selected on the basis of their high and low feed conversion ratio. Microbial DNA was extracted from rumen samples, and the relative abundances of microbial genes were determined *via* shotgun metagenomic sequencing. Using partial least squares analyses, we identified sets of 20, 14, 17, and 18 microbial genes whose relative abundances explained 63, 65, 66, and 73% of the variation of feed conversion efficiency, average daily weight gain, residual feed intake, and daily feed intake, respectively. The microbial genes associated with each of these traits were mostly different, but highly correlated traits such as feed conversion ratio and growth rate showed some overlapping genes. Consistent with this result, distinct clusters of a coabundance network were enriched with microbial genes identified to be related with feed conversion ratio and growth rate or daily feed intake and residual feed intake. Microbial genes encoding for proteins related to cell wall biosynthesis, hemicellulose, and cellulose degradation and host–microbiome crosstalk (e.g., *aguA, ptb*, K01188, and *murD*) were associated with feed conversion ratio and/or average daily gain. Genes related to vitamin B12 biosynthesis, environmental information processing, and bacterial mobility (e.g., *cobD*, *tolC*, and *fliN*) were associated with residual feed intake and/or daily feed intake. This research highlights the association of the microbiome with feed conversion processes, influencing growth rate and appetite, and it emphasizes the opportunity to use relative abundances of microbial genes in the prediction of these performance traits, with potential implementation in animal breeding programs and dietary interventions.

Keywords: feed conversion efficiency, appetite, metagenomics, rumen microbiome, microbial gene networks

INTRODUCTION

The global population is expected to reach 9.8 billion by 2050 (United Nations–Department of Economic and Social Affairs/Population Division, 2017), resulting in an escalation of the global demand for food and of the need for economically and environmentally sustainable livestock production systems (Godfray et al., 2010; Gerber et al., 2013). A large portion of livestock production is based on ruminants. In 2017, the EU-28 had a population of 88 million bovine animals, including cattle and water buffalo (Eurostat, 2018). Ruminants are particularly interesting due to their ability to convert human-indigestible plant biomass into high-quality products for human consumption such as meat and milk. Ruminants live in a symbiotic relationship with their rumen microbiota (comprising bacteria, protozoa, fungi, and archaea), which produce enzymes able to digest their food by breaking down complex polysaccharides of the plant biomass into volatile fatty acids (VFA), microbial proteins, and vitamins (Russell and Hespell, 1981; Bergman, 1990; Van Soest, 1994). Thus, the rumen microbiota fermentation profile has a significant influence on the feed conversion efficiency of the host (Russell, 2001; Li et al., 2009; Hernandez-Sanabria et al., 2011; Jami et al., 2014; Sasson et al., 2017; Meale et al., 2018) and is accountable for up to 70% of the host's daily energy requirements (Bergman, 1990).

In beef cattle production systems, expenses associated with feed account for up to 75% of the total production costs (Moran, 2005a; Nielsen et al., 2013), which makes the improvement of feed conversion efficiency very economically compelling. There is consequently great interest in understanding the host–microbial symbiotic relationships responsible for the conversion of feed into energy, protein, and vitamins usable by the host animal, but the mechanisms and degree to which the rumen microbiome impacts on animal production, health, and efficiency remain undercharacterized (Brulc et al., 2009; Creevey et al., 2014). Although the rumen harbors a core microbiome (Jami and Mizrahi, 2012; Henderson et al., 2015), in agreement with studies performed in the human gastrointestinal tract (Tap et al., 2009; Qin et al., 2010), the structure, and composition of the rumen microbiome varies within and between animals with differing performance traits. For example, in lactating dairy cattle, the increased methane yield during late lactation in comparison to early lactation within the same individual was found to be associated with significant changes in the ruminal microbial community structure (Lyons et al., 2018); Myer et al. (2015) showed different relative abundances of some microbial taxa and operational taxonomic units in animals with different average daily gain (ADG); Shabat et al. (2016) focused on residual feed intake (RFI) to demonstrate that highly efficient animals had a less diverse microbiota, being dominated by specific taxa and microbial genes which were involved in simpler metabolic pathway networks when compared to their less efficient counterparts. Other authors have reported that the rumen microbiome varies more between animals than within animals, proposing that the host itself and its physiological parameters have a significant influence on its own rumen microbiome (Li et al., 2009) and, therefore, on the efficiency of feed conversion into energy. In a mouse study, Benson et al. (2010) found that there is a well-defined portion of the gut microbiota that is subject to host genetic control, proposing it to be regarded as a host trait, rather than an environmental trait affecting the host. In agreement, in a beef cattle study, Roehe et al. (2016) confirmed the host genetic influence on the rumen bacterial composition using a genetic model based on sire progeny groups. The differences between sire progeny groups in methane emissions were in some cases larger than the differences found between diets differing largely in plant fiber content, suggesting a substantial host genetic influence on the microbial communities.

Selecting animals for breeding based on their ability to harvest energy from feed, together with nutritional interventions, could be the basis for an effective strategy to produce faster growing and more efficient animals (Gerber et al., 2013; Scollan et al., 2018). Given that the host has influence over the ruminal microbiome, which impacts the animals' feed conversion efficiency, this selection may be further improved by the inclusion of rumen metagenomic information into predictive models, as previously suggested by Ross et al. (2013). Feed conversion efficiency is very often estimated by either feed conversion ratio (FCR) or RFI; the latter is independent of growth and maturity patterns and is expected to be more sensitive and precise in measurements of feed utilization (Arthur and Herd, 2008). The use of microbial genes as proxies for feed conversion efficiency traits may be much more cost effective, rapid, and less labor intensive than their recording (Ross et al., 2013; Roehe et al., 2016). Our earlier research was the first proposing that the inclusion of relative abundance of microbial genes as proxies for FCR may be favorable, allowing their use as selection criteria for breeding animals, by identifying 49 microbial genes that explained 88.3% of the variation observed in FCR (Roehe et al., 2016). To our knowledge, no other studies have focused on the relationship between microbial gene abundances and RFI, daily feed intake (DFI), and ADG, which highlights the importance and novelty of the present work.

This study aimed at validating whether rumen microbial gene abundances are suitable proxies for feed conversion efficiency traits such as FCR; the analysis was further extended by focusing on RFI. Based on the previous evidence of strong interactions between the rumen microbiome and the host animal with consequences for feed conversion efficiency (Guan et al., 2008; Roehe et al., 2016; Shabat et al., 2016), we hypothesized that microbial gene abundances are linked to the animals' appetite and, consequently, to feed intake. A further aim of this research was to gain insight into the association of growth rate with the microbial gene abundances. Building on this, we aimed at better understanding the rumen microbial functional network associated with feed conversion efficiency and its component traits. This research will improve on the current knowledge about the impact of the rumen microbiome on appetite, growth, and efficiency of feed conversion processes.

MATERIALS AND METHODS

Ethics Statement

This study was conducted at the Beef and Sheep Research Centre, SRUC, UK. The study was carried out in accordance with the requirements of the UK Animals (Scientific Procedures) Act 1986. The protocol was approved by the Animal Experiment Committee of SRUC. All standard biosecurity and institutional

safety procedures were applied during the animal experiment and the laboratory analysis.

Animals, Adaptation Period, and Measurement of Traits

Two experiments were carried out to determine the effect of nitrate or lipid additives within different basal diets on methane emissions from beef cattle. The first experiment was conducted in 2013, and it consisted of a 2 × 2 × 3 factorial design including 84 steers of two breed types (crossbreed Charolais, CHx and Luing); two basal diets, forage (FOR) and concentrate (CONC), which consisted respectively of ratios of 520:480 and 84:916 forage to concentrate (g/kg dry matter); and three treatments, nitrate and lipid feed additives, as well as the control. From these animals, 24 animals were selected with extreme high and low FCR values within breed type and basal diet (two animals per feed additive and control). More details related to this experiment can be found in Duthie et al. (2015) and Troy et al. (2015). The second experiment was a 2 × 4 factorial design experiment, conducted in 2014, involving 80 animals. There were two breed types—40 crossbred Limousin (LIMx) and 40 crossbred Aberdeen Angus (AAx)—which were subject to a balanced design consisting of four dietary treatments using one basal diet (550:450 forage to concentrate ratio g/kg dry matter, FOR) and testing the effects of feed additives nitrate, lipid, or their combination in comparison to the control on methane output. Full details of the experiment are presented in Duthie et al. (2017). From this experiment, 18 animals were selected within each combination of breed type and diet: nine for the high FCR group and nine for the low FCR group. DFI was assessed by measuring dry matter intake (DMI, kg/day), which was recorded in both experiments using electronic feeding equipment (HOKO, Insentec, Marknesse, The Netherlands). Body weight (BW) was measured weekly using a calibrated weight scale (before fresh feed was offered). Growth was modeled by linear regression of BW against test date to obtain ADG, mid-test BW, and mid-test metabolic BW (MBW = $BW^{0.75}$). FCR was calculated as average DMI (kg/day) divided by ADG. RFI was estimated as deviation of actual DMI (kg/day) from DMI predicted based on linear regression of actual DMI on ADG, mid-MBW, and fat depth at $12^{th}/13^{th}$ rib at the end of the 56-day test (Duthie et al., 2015; Troy et al., 2015; Duthie et al., 2017).

A flowchart summarizing the methods for generation of data and subsequent statistical analyses is presented in **Figure 1**.

Sampling of Rumen Digesta and Whole Metagenomic Sequencing

As described in Duthie et al. (2015) and Auffret et al. (2017), animals from both experiments were slaughtered in a commercial

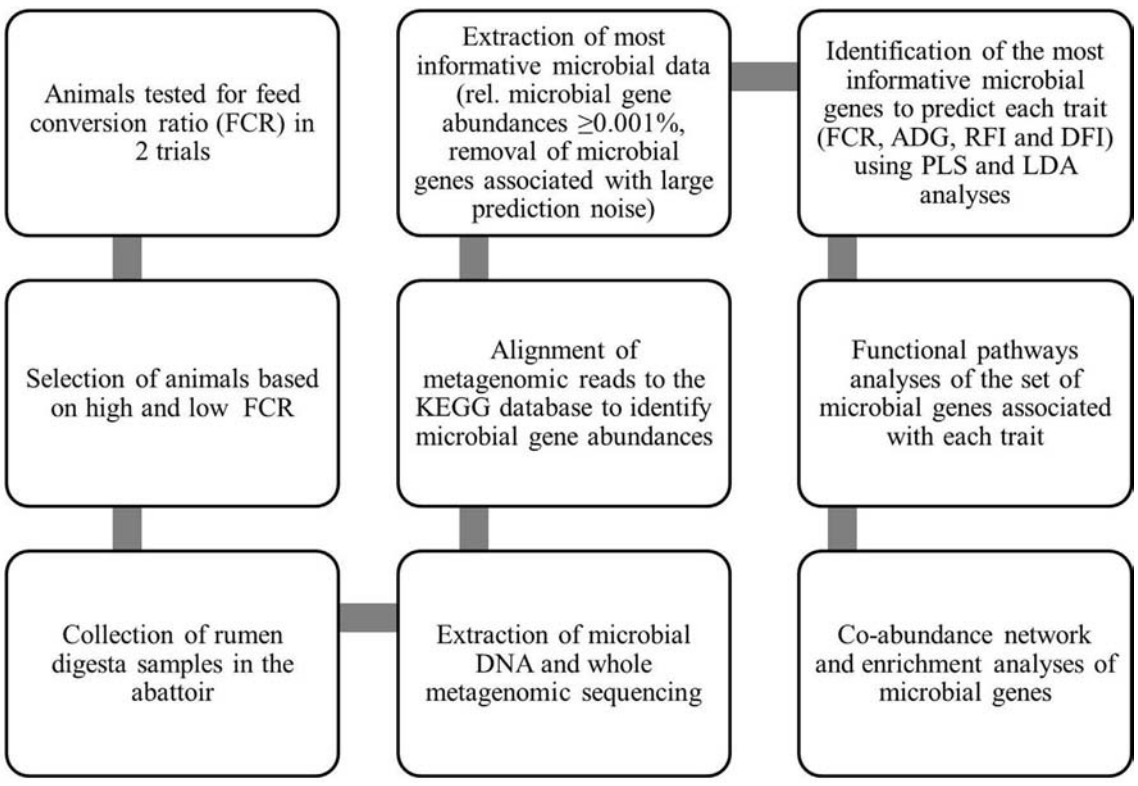


FIGURE 1 | Flowchart summarizing methods for generation of data and their statistical analyses: This flowchart summarizes how the data were generated and which statistical analyses were used to identify the associations between gene abundances and performance traits of animals to understand the rumen microbial functional pathways associated with these traits. KEGG, Kyoto Encyclopedia of Genes and Genomes; FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; PLS, partial least squares; LDA, linear discriminant analysis.

abattoir where two samples of rumen digesta (~50 ml) were collected immediately after the rumen was opened to be drained. The slaughter house sample collection process results in well-mixed samples of rumen contents. DNA was extracted from the rumen samples of 42 animals following the methodology described in Rooke et al. (2014). Illumina TruSeq libraries were prepared from genomic DNA and sequenced on Illumina HiSeq systems 4000 by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 8 and 15 GB per sample (between 40 and 73 million paired reads). The raw data can be downloaded from the European Nucleotide Archive under accession PRJEB21624.

Identification of the Rumen Microbial Gene Abundances

Bioinformatics analysis for identification of rumen microbial genes was carried out as previously described by Wallace et al. (2015). Briefly, to measure the abundance of known functional microbial genes in the rumen samples, reads from whole metagenome sequencing were aligned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) using Novoalign (www.novocraft.com). Parameters were adjusted such that all hits were reported that were equal in quality to the best hit for each read and allowing up to a 10% mismatch across the fragment. The KEGG Orthologue groups (KO) of all hits that were equal to the best hit were examined. If we were unable to resolve the read to a single KO, the read was ignored; otherwise, the read was assigned to the unique KO. Read counts were summed and normalized to the total number of hits. This mapping of the whole metagenomic data to the KEGG database resulted in a dataset comprising of 4,966 KEGG genes. Microbial genes were removed from the dataset when they were absent from three or more animals and when the mean relative abundance was lower than 0.001%, leaving 1,692 microbial genes for further analyses.

Statistical Analysis

For each of the 1,692 microbial genes, a linear model was fitted, including as fixed effects a combined class variable of breed, diet, and year of experiment (six levels) and the FCR groups (high FCR, FCR-H and low FCR, FCR-L) using the lm() function in R version 3.4.2. The microbial genes which resulted in $P \geq 0.1$ for the differences in FCR groups were not considered in the partial least squares analyses (PLS, SAS version 9.3 for Windows, SAS Institute Inc., Cary, NC, USA) to avoid excessive noise of microbial genes uncorrelated to the traits of interest. In the linear model, FCR groups were replaced successively by ADG, RFI, and DFI as covariables to identify only potentially relevant microbial genes of these traits for further PLS analyses. In addition, genes with unknown function were removed from these datasets.

Microbial genes whose relative abundances were significantly associated to each trait in the linear models were analyzed using a sequential PLS-based methodology. First, PLS models were calculated in which the number of latent variables was determined by "leave-one-out" cross-validation, and genes with lower variable importance in projection (VIP) were removed. Second, the sets of genes created in the first step were evaluated by PLS models using three latent variables to determine the smaller set of genes leading to higher explained variation of both independent and dependent variables.

Each set of microbial genes identified in the PLS analyses as best predicting the trait was then used in a linear discriminant analysis (LDA), performed in R version 3.5.1 (2018-07-02) package MASS_7.3-51.4. In these analyses, the categories were for FCR those described previously as FCR-H and FCR-L; for all other traits, animals were classified as high or low, depending on their observations being higher or lower than the median (balanced for trial, breed, and diet).

The microbial genes identified to be significantly associated with each trait were submitted to an extensive review about their functionality based on databases such as KEGG (Kanehisa Laboratories, 2018), BioCyc (Karp et al., 2017), and UniProt (Bateman, 2019) and information from the literature.

Networks

The coabundances between microbial genes were investigated in a stepwise network analysis using the Graphia Professional software (Kajeka Ltd, Edinburgh; Freeman et al., 2007), in which nodes represent microbial genes and edges represent a correlation value above a defined value of *r*. In the first step, the correlation threshold of $r = 0.45$ was selected such that all microbial genes ($n = 1{,}692$) were included in the network. The microbial genes identified by PLS to be associated with a trait of interest were then located in the network. Clustering was performed using the Markov clustering method (MCL) available in Graphia Professional using the default settings (inflation, preinflation, and scheme values of 6). All clusters that held at least one microbial gene previously identified in the PLS analysis to be associated with a trait of interest were identified. These were incorporated into a new network generated at correlation threshold of $r = 0.80$ containing 1,135 microbial genes. MCL was then performed on this network, with inflation and preinflation values of 2 and scheme value of 6, reflecting the clustering structure suggested in the network itself. Analyses of enrichment of genes identified in the PLS as associated to each trait were performed on the clusters, and significance was assessed at $P < 0.05$.

RESULTS

Performance Traits Related With Feed Conversion Efficiency

The average FCR values observed for animals selected into FCR-H (inefficient) and FCR-L (efficient) groups differed significantly by 2.3 kg DFI/kg ADG (**Figure 2**). When comparing these two groups for other traits, the FCR-H group had significantly higher values of RFI (0.8 kg) and significantly lower ADG (0.39 kg); in the case of DFI, no significant difference was observed between the FCR groups.

ADG and FCR had a strong significant negative correlation of 0.80, suggesting that high growth rate is associated with efficient animals, using less feed per kilogram of weight gain. FCR and RFI were significantly positively correlated, but at a low level of

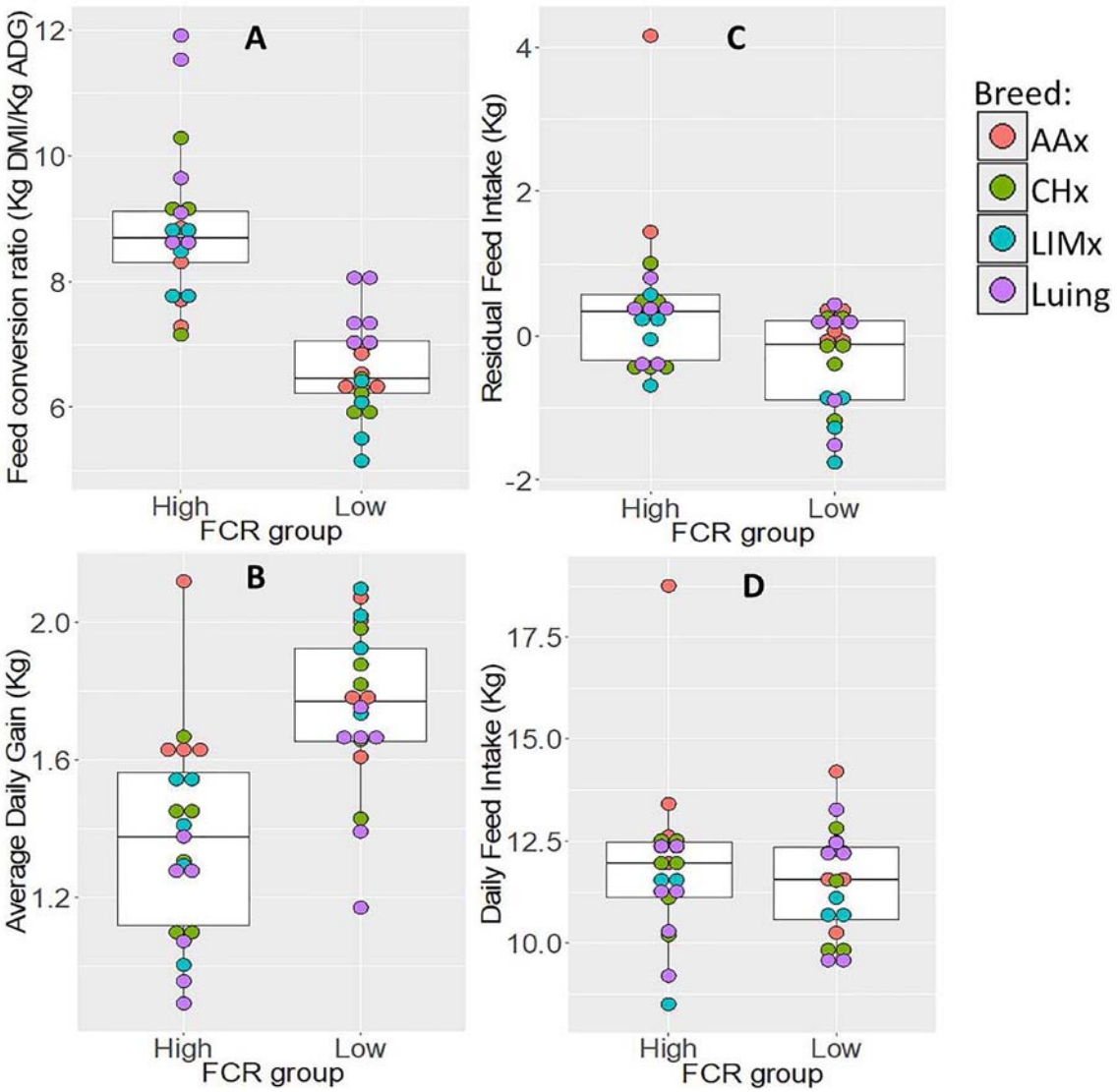


FIGURE 2 | Distribution of variation and range of performance traits: **(A)** feed conversion ratio, **(B)** average daily gain, **(C)** residual feed intake, and **(D)** daily feed intake within feed conversion ratio groups (high and low). The boxplots show the variation and range of each trait within each feed conversion ratio group. FCR, feed conversion ratio; AAx, crossbred Aberdeen Angus; CHx, crossbred Charolais; LIMx, crossbred Limousin.

0.32. DFI was significantly correlated with RFI and ADG at high and moderate levels of 0.77 and 0.53, respectively.

Rumen Microbial Genes Associated With Feed Conversion Efficiency Traits

The PLS analyses identified sets of 20 and 14 microbial genes whose relative abundances explained 63.4 and 65.4% of the variation in FCR and ADG, respectively, and sets of 17 and 18 microbial genes whose relative abundances explained 65.6 and 72.9% of the variation in RFI and DFI, respectively, including the combined fixed effect of diet, breed, and year of experiment (**Table 1**). Without this combined fixed effect, the variances explained by microbial genes in FCR and ADG decreased to 54.2 and 61.4%, while in RFI and DFI, they decreased to 50.8 and 67.7%, respectively. A discriminant analysis between groups of high- and low-performing animals, using the set of microbial genes identified in the PLS analysis to best predict each trait, resulted in prediction accuracies of 90, 79, 86, and 86% for FCR, ADG, RFI, and DFI (**Figure 3**).

The Venn diagram presented in **Figure 4** illustrates the overlap between the sets of genes identified for the prediction of each of the four traits. For the prediction of FCR and ADG, six microbial genes were simultaneously selected: UDP-N-acetylmuramoylalanine-D-glutamate ligase, glycine cleavage system H protein, translation initiation factor IF-1, N utilization substance protein A, DNA-binding protein HU-beta, and diphthamide synthase subunit *dph2* (*murD*, *gcvH*, *infA*, *nusA*, *hupB*, and *dph2*, respectively). Three microbial genes were

TABLE 1 | Percentage of variation in each trait explained by the microbial genes identified in the partial least squares (PLS).

		Percent variation accounted for by partial least squares factors			
		Model effects		Dependent variables	
Trait	No. factors	Current	Total	Current	Total
FCR	1	41.59	41.59	35.46	35.46
	2	6.35	47.94	21.19	56.65
	3	7.57	55.51	6.72	63.37
ADG	1	39.42	39.42	49.26	49.26
	2	9.60	49.02	11.47	60.73
	3	7.97	56.99	4.67	65.40
RFI	1	24.04	24.04	44.32	44.32
	2	13.95	37.99	16.80	61.12
	3	16.72	54.71	4.52	65.63
DFI	1	28.98	28.98	44.94	44.94
	2	21.25	50.23	19.94	64.88
	3	7.86	58.09	8.05	72.93

The number of factors refers to the number of latent variables in which the total number of microbial genes (independent variables) were projected in the PLS procedure, and each factor accounts for a portion of the total explained variation. The "Model Effects" columns refer to the percent variability of the independent variables matrix that relates to the respective percent variability presented in the "Dependent Variables" columns. The "Current" columns present values for each extracted factor individually, and the "Total" columns present the subtotal variation. The cells colored in gray contain the values of percent variation explained by the three latent variables for each trait. FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake.

simultaneously selected for the prediction of traits RFI and DFI: glucose-1-phosphate cytidylyltransferase, CDP-glucose 4,6-dehydratase, and energy-converting hydrogenase B subunit D (*rfbF*, *rfbG*, and *ehbD*, respectively). The microbial genes identified for the prediction of more than one trait are highlighted in the shaded rows in **Tables 2–5**, in which a more detailed information about their function and importance for prediction is provided.

Based on the relative abundance of 1,135 microbial genes across rumen samples, a coabundance network was developed (**Figure 5**), and clusters were identified. The clustering pattern evidences the microbial genes that are more closely connected to microbial genes previously identified in the PLS analyses. The network cluster to which each microbial gene belongs to is presented in **Tables 2–5**. Cluster 2 was significantly enriched for microbial genes predicting DFI and RFI&DFI (RFI and/or DFI). Cluster 4 was enriched for microbial genes predicting RFI and RFI&DFI. Microbial genes simultaneously predicting FCR and ADG were enriched in clusters 20 and 21, while those predicting FCR&ADG (FCR and/or ADG) were enriched in clusters 21 and 25. ADG-predicting microbial genes were enriched in clusters 21 and 25, whereas FCR-predicting genes were only enriched in cluster 25. Other genes previously identified in the PLS analysis were scattered across the graph.

Most microbial genes identified exclusively for the prediction of FCR are related to carbohydrate metabolism and transport: fructuronate reductase, galactokinase, alpha-glucuronidase, beta-glucuronidase, beta-glucosidase, phosphate butyryltransferase P, UDP-N-acetylglucosamine acyltransferase, gluconate 5-dehydrogenase, and lactate permease (respectively *uxuB*, *galK*, *aguA*, *uidA*, K01188, *ptb*, *lpxA*, *idnO*, and *lctP*) were proportionally more abundant in efficient animals (lower FCR, **Supplementary Figure S1A**). The microbial gene lactoylglutathione lyase (*glo1*) is also associated with carbohydrate metabolism and identified for predicting FCR, but it had higher relative abundance in less efficient animals (higher FCR). Microbial genes *galK* and *xylE* (i.e., MFS transporter, SP family, xylose:H+ symporter) were both located in cluster 5, but this cluster was not significantly enriched for microbial genes associated to FCR. On the other hand, cluster 25 was enriched due to the presence of microbial genes *uxuB* and *lpxA*.

Microbial genes associated with amino acid metabolism and transport pathways were identified for the prediction of ADG and found to be relatively more abundant in animals with higher ADG (see **Supplementary Figure S1B**), e.g., aspartate-semialdehyde dehydrogenase and phenylacetate-CoA ligase (*asd* and *paak*, respectively). Some housekeeping genes were also identified for this set, including large subunit ribosomal protein L17 and L36, F-type H+-transporting ATPase subunit delta and FKBP-type peptidyl-prolyl cis-trans isomerase *slyD* (*rplQ* and *rpmJ*, *atpH*, and *slyD*). Genes *rplQ*, *atpH*, and *slyD* were relatively more abundant in animals with higher ADG, and *rpmJ* was relatively more abundant in animals with lower ADG. The microbial gene N-acetylmuramoyl-L-alanine amidase (*amiABC*) was identified for prediction of ADG, being negatively correlated with the trait.

All microbial genes simultaneously identified for predicting FCR and ADG showed a negative correlation to FCR and a positive correlation to ADG. These included housekeeping genes (*infA*, *hupB*, and *dph2*), a gene related to carbohydrate metabolism (*gcvH*), *murD*, which was associated with peptidoglycan metabolism and D-glutamine and D-glutamate metabolism, and *nusA*, associated with transcription regulation. Cluster 21 was enriched in ADG- and FCR&ADG-predicting microbial genes due to the presence of *atpH*, *rplQ* (ADG), and *infA* (FCR&ADG).

Five microbial genes identified for the prediction of RFI were associated with environmental sensing, bacterial chemotaxis, and motility: sensor kinase *cheA*, response regulator *cheY*, methyl accepting chemotaxis protein, flagellar motor switch protein *fliN/fliY*, and flagellar hook protein *flgE* (*cheA*, *cheY*, *mcp*, *fliN*, and *flgE*, respectively) were found to be relatively more abundant in more efficient animals, i.e., lower RFI. Other microbial genes associated with RFI are involved in the biosynthesis of cofactors and vitamins, particularly vitamin B12 production, for example, cobalt transport protein, threonine-phosphate decarboxylase, and precorrin-6Y C5,15-methyltransferase (decarboxylating), which correspond respectively to *cbiN*, *cobD*, and *cobL* (**Supplementary Figure S1C**). Finally, three genes that encode proteins related to carbohydrate transport and metabolism were relatively more abundant in more efficient animals (i.e., lower RFI): the simple sugar transport system permease protein, oxaloacetate decarboxylase, alpha subunit, and aldehyde:ferredoxin oxidoreductase (respectively ABC.

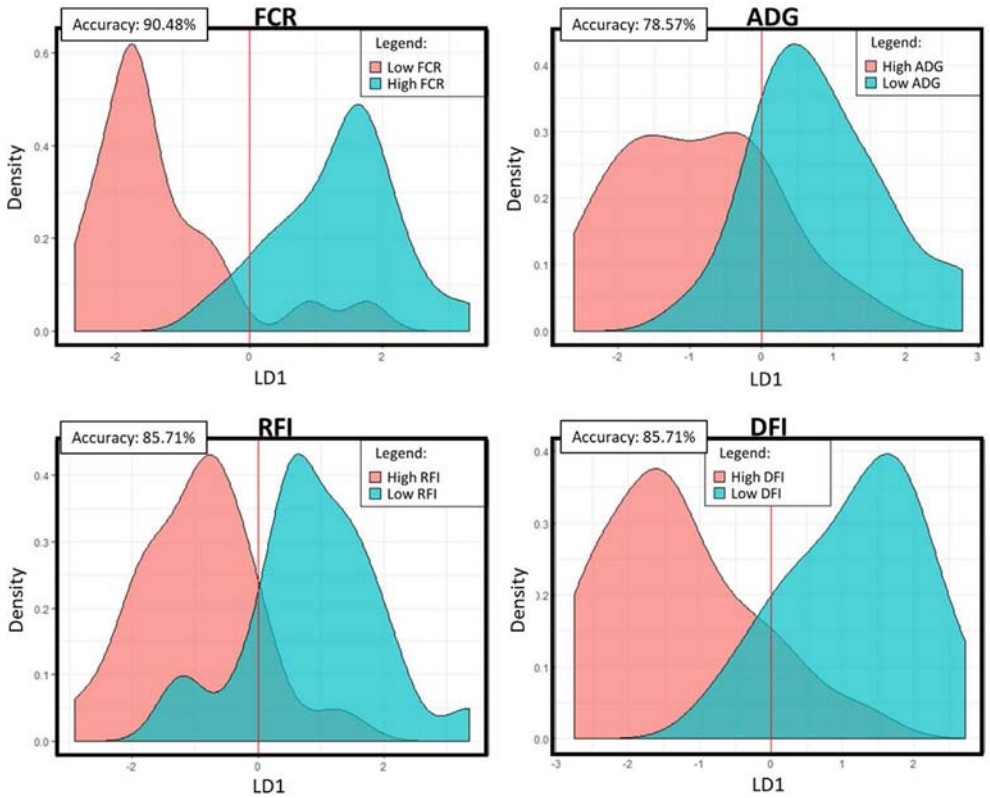


FIGURE 3 | Linear discriminant analysis density plots: Microbial genes identified in the PLS analyses to be significantly associated with the trait were used in a linear discriminant analysis of high- and low-performing animals. The density plots represent the predicted categories for each trait. The accuracy value represents the percentage of animals that were correctly assigned to their category. FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; LD1, linear discriminant 1.

SS.P, *oadA*, and *aor*). Cluster 4 was significantly enriched in microbial genes associated with RFI due to the presence of microbial genes *cobD*, *cobL*, *mcp*, and *oadA*, and serine-type D-Ala-D-Ala carboxypeptidase (penicillin-binding protein 5/6), inner membrane protein, and Cd2+/Zn2+-exporting ATPase (respectively, *dacC*, *ybrG*, and *zntA*).

The set of microbial genes identified for prediction of DFI included four microbial genes, proportionally more abundant in animals with higher DFI, which encoded proteins associated with environmental sensing, i.e., nitrogen regulatory protein P-II 1, outer membrane channel protein *TolC*, and preprotein translocase subunit *YajC* (*glnB*, *tolC*, and *yajC*, respectively). Nitrate reductase 1, alpha subunit (*narG*) was related to denitrification, releasing nitrite, and it was found to be relatively more abundant in animals with lower DFI (**Supplementary Figure S1D**). DNA-directed RNA polymerase subunit beta (*rpoB*, proportional higher abundance in animals with lower DFI), ribosomal large subunit pseudouridine synthase B, exodeoxyribonuclease VII small subunit, ribonuclease III, N utilization substance protein B, and integration host factor subunit alpha (respectively *rluB*, *xseB*, *rnc*, *nusB*, and *ihfA*, proportionally more abundant in animals with higher DFI) are housekeeping genes identified in this work for the prediction of DFI. Cluster 2 was significantly enriched with microbial genes associated with DFI due to the presence of *glnB*, *infA*, *mrdA*, *nusB*, *rdgB*, *rluB*, *tolC*, and *xseB*.

RFI- and DFI-predicting genes include glucose-1-phosphate cytidylyltransferase, CDP-glucose 4,6-dehydratase (respectively *rfbF* and *rfbG*, related to amino sugar and nucleotide sugar metabolism), and energy-converting hydrogenase B subunit D (*ehbD*, housekeeping). These three genes were proportionally more abundant in less efficient animals (higher RFI associated with increased DFI).

DISCUSSION

Rumen Microbial Gene Abundances Associated With Efficiency Traits

Our research indicates that there is a substantial link between rumen microbial gene abundances and appetite (measured as feed intake), growth rate, and feed conversion efficiency (**Figure 6**). The relative abundances of 20 and 17 microbial genes accounted for substantial variation (>60%) in FCR and RFI, respectively. The discriminant analyses of high- and low-performing animals indicated that accurate classification (>85% correct assignment of FCR and RFI categories) could be achieved using the microbial genes identified in the PLS for the prediction of the traits. Roehe et al. (2016) also found an association of microbial gene abundances with FCR, but their results were based on a smaller number of animals selected for their extreme values in

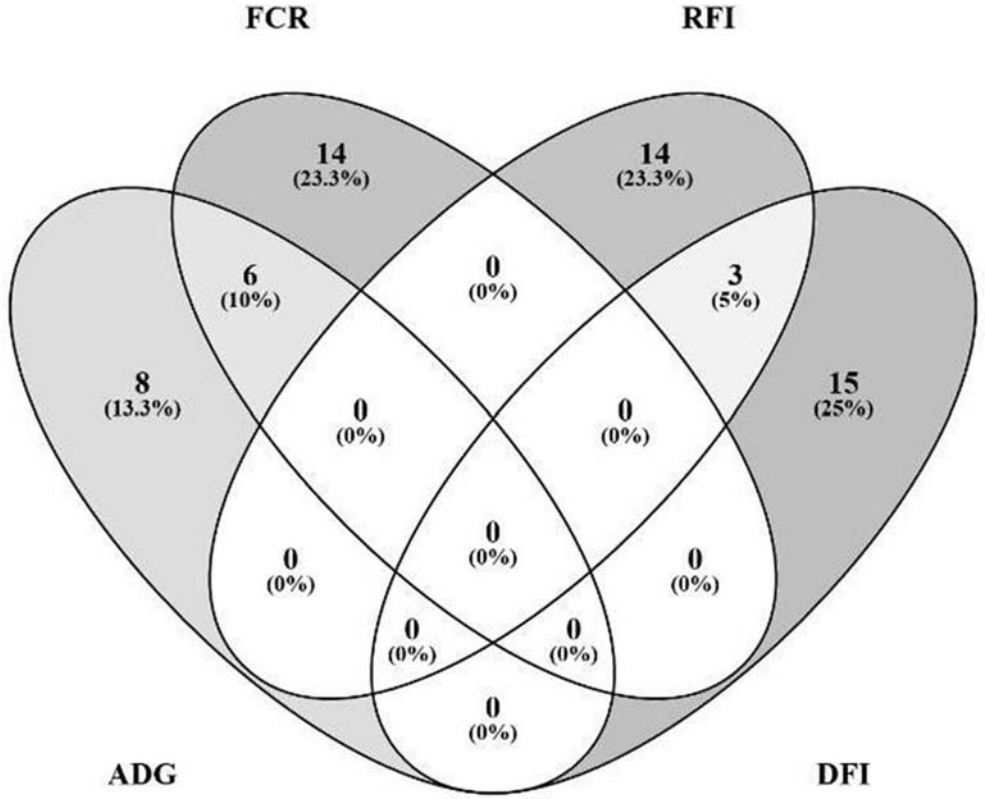


FIGURE 4 | Overlap analysis of identified microbial genes: The image illustrates the number of microbial genes identified in the partial least squares analysis as fitted for prediction of each animal performance trait exclusively, and the number of microbial genes simultaneously predicted for multiple traits: six microbial genes were simultaneously identified for prediction of FCR (feed conversion ratio) and ADG (average daily gain), and three for both RFI (residual feed intake) and DFI (daily feed intake).

methane emissions. In the present study, animals were selected based on their extreme FCR values, yielding a statistically more powerful estimate of this trait. Whereas FCR is calculated as a ratio between DFI and ADG and is therefore highly affected by growth rate and body composition, RFI is independent of these traits (Berry and Crowley, 2013). The low phenotypic correlation ($r = 0.32$) between FCR and RFI suggests that these traits capture substantially distinct characteristics.

For ADG and DFI, the relative abundances of 14 and 18 microbial genes, respectively, also explained substantial variation (>65%), and the discriminant analyses of high- and low-performing animals resulted in high prediction accuracies of 79 and 86%, respectively. These component traits were moderately correlated, agreeing with the report by Berry and Crowley (2013) of a large independent variation of feed intake and weight gain.

The animals' appetite, feeding behaviour, and gastrointestinal motility (among other traits) are thought to be regulated by several mechanisms, including a communication between the rumen microbiome and the brain, through the gut–liver–brain axis (vagus nerve). This communication has been proposed to be mediated by multiple mechanisms, such as insulin/glucagon homeostasis, oxidation of acetyl coenzyme A, and release of VFA by the rumen microbiota (like propionate, associated with hypophagic behavior in ruminants, or butyrate and acetate, associated with motility of the gastrointestinal tract in monogastric animals; Sakata and Tamate, 1979; Cherbut, 2003; Oba and Allen, 2003; Arora et al., 2011; Maldini and Allen, 2018). Given the predictability of performance traits using relative abundances of rumen microbial genes observed in the present research (particularly that of DFI) and the high impact of the rumen microbiome on feed intake regulation (as discussed in the literature), we hypothesize that rumen microbial genes are closely involved in the metabolic pathways that regulate feed intake.

Differential Microbial Gene Sets Predicting Distinct Trait Complexes

The coabundance microbial gene network (**Figure 5**) identified two separate trait complexes. While microbial genes identified for the prediction of FCR were grouped with ADG-predicting genes, microbial genes identified for the prediction of RFI were grouped with DFI-predicting genes, as revealed by differential enrichment in separate clusters (**Supplementary Figure S2**). For example, beta-glucosidase is encoded by microbial genes *bglX* and K01188, which were associated to different traits (DFI and FCR, respectively). This type of differential clustering was previously observed for microbial genes associated with methane

TABLE 2 | Summary of microbial genes identified for the prediction of FCR.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K03783	Purine-nucleoside phosphorylase	*punA*	[2]Metabolic pathways; biosynthesis of secondary metabolites; purine metabolism; pyrimidine metabolism; nicotinate and nicotinamide metabolism	0.0107	–0.2755	1.22	[1]
K08138	MFS transporter, SP family, xylose:H+ symporter	*xylE*	[3]Carbohydrate transport and metabolism, amino acid transport and metabolism, Inorganic ion transport and metabolism	0.0404	0.1135	1.09	05
K00046	Gluconate 5-dehydrogenase	*idnO*	[4]L-idonate degradation	0.0845	0.0166	1.08	11
K00040	Fructuronate reductase	*uxuB*	[2]Metabolic pathways; pentose and glucuronate interconversions	0.0847	0.0503	1.01	25
K01759	Lactoylglutathione lyase	*glo1*	[2]Pyruvate metabolism	0.0021	0.1547	1.01	09
K00849	Galactokinase	*galK*	[2]Metabolic pathways; galactose metabolism; amino sugar and nucleotide sugar metabolism	0.0631	0.0675	1.00	05
K01195	Beta-glucuronidase	*uidA*	[2]Metabolic pathways; biosynthesis of secondary metabolites; pentose and glucuronate interconversions; glycosaminoglycan degradation; porphyrin and chlorophyll metabolism; flavone and flavonol biosynthesis; drug metabolism—other enzymes; lysosome	0.0127	–0.1174	0.99	07
K14220	tRNA Asn	*tRNA-Asn*	[2]Aminoacyl-tRNA biosynthesis	0.0155	0.0139	0.96	NC
K00677	UDP-N-acetylglucosamine acyltransferase	*lpxA*	[2]Metabolic pathways; lipopolysaccharide biosynthesis; cationic antimicrobial peptide (CAMP) resistance	0.0403	–0.0186	0.91	25
K01188	Beta-glucosidase	beta-glucosidase	[2]Metabolic pathways; biosynthesis of secondary metabolites; cyanoamino acid metabolism; starch and sucrose metabolism; phenylpropanoid biosynthesis	0.0398	–0.0210	0.90	[1]
K07214	Enterochelin esterase and related enzymes	*fes*	[3]Inorganic ion transport and metabolism	0.0475	–0.1511	0.90	[1]
K03303	Lactate permease	*lctP*	[5]Lactate transmembrane transporter activity	0.0195	–0.0271	0.88	28
K00634	Phosphate butyryltransferase	*ptb*	[2]Metabolic pathways; butanoate metabolism	0.0075	–0.0079	0.85	[1]
K01235	Alpha-glucuronidase	*aguA*	[3]Carbohydrate transport and metabolism	0.0104	–0.0626	0.80	NC
K07561	diphthamide synthase subunit DPH2	*dph2*	[3]Translation, ribosomal structure and biogenesis	0.0030	–0.3881	1.86	01
K01925	UDP-N-acetylmuramoylalanine–D-glutamate ligase	*murD*	[2]Metabolic pathways; D-Glutamine and D-glutamate metabolism; peptidoglycan biosynthesis	0.0620	–0.0857	0.99	[1]
K02437	Glycine cleavage system H protein	*gcvH*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0069	–0.0167	0.93	[1]
K03530	DNA-binding protein HU-beta	*hupB*	[3]DNA binding protein: replication, recombination, and repair	0.0331	–0.0892	0.89	19
K02600	N utilization substance protein A	*nusA*	[3]Transcription	0.1126	–0.0655	0.89	[1]
K02518	Translation initiation factor IF-1	*infA*	[3]Translation, ribosomal structure and biogenesis	0.0346	0.0170	0.88	21

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 63.4% of the variation in FCR (feed conversion ratio). Rows colored in gray correspond to genes simultaneously identified for both FCR and ADG (average daily gain) prediction.

emissions and FCR by Roehe et al. (2016). The trait complexes associated with feed conversion efficiency were further evidenced when analyzing the overlapping genes identified for the prediction of each trait (**Figure 4** and shaded rows in **Tables 2**–**5**), i.e., six microbial genes were identified for the prediction of both FCR and ADG and three genes for the prediction of both RFI and DFI. In agreement, strong correlations were observed for each pair of traits, as shown previously in the literature with the literature (Arthur and Herd, 2008; Herd et al., 2014). These results suggest that different microbial genes can be used to predict each trait. Furthermore, microbial genes overlapping for the prediction of more than one trait might be useful for the interpretation of biological processes explaining the correlation between phenotypes.

TABLE 3 | Summary of microbial genes identified for the prediction of ADG.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K01448	N-acetylmuramoyl-L-alanine amidase	*amiABC*	[2]Cationic antimicrobial peptide (CAMP) resistance	0.0236	–0.1937	1.22	06
K00133	Aspartate-semialdehyde dehydrogenase	*asd*	[2]Metabolic pathways; microbial metabolism in diverse environments; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; monobactam biosynthesis; cysteine and methionine metabolism; lysine biosynthesis; 2-oxocarboxylic acid metabolism; biosynthesis of amino acids	0.1197	–0.0684	1.20	NC
K01912	Phenylacetate-CoA ligase	*paaK*	[2]Microbial metabolism in diverse environments; phenylalanine metabolism; biofilm formation—*Vibrio cholerae*	0.1543	–0.0980	1.16	16
K02919	Large subunit ribosomal protein L36	*rpmJ*	[2]Ribosome	0.0261	–0.1884	1.04	NC
K02879	Large subunit ribosomal protein L17	*rplQ*	[2]Ribosome	0.0773	0.0746	1.00	21
K02113	F-type H+-transporting ATPase subunit delta	*atpH*	[2]Metabolic pathways; oxidative phosphorylation; photosynthesis	0.0292	–0.0486	1.00	21
K00283	Glycine dehydrogenase subunit 2	*gcvPB*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0284	0.0502	0.99	25
K03775	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD	*slyD*	[5]Posttranslational modification, protein turnover, chaperones	0.0139	0.0672	0.93	22
K07561	Diphthamide synthase subunit DPH2	*dph2*	[5]Translation, ribosomal structure, and biogenesis	0.0030	0.2310	1.20	01
K01925	UDP-N-acetylmuramoylalanine–D-glutamate ligase	*murD*	[2]Metabolic pathways; D-glutamine and D-glutamate metabolism; peptidoglycan biosynthesis	0.0620	0.1155	1.15	[1]
K02437	Glycine cleavage system H protein	*gcvH*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0069	0.1209	1.08	[1]
K03530	DNA-binding protein HU-beta	*hupB*	[5]DNA binding protein: replication, recombination, and repair	0.0331	0.1062	1.07	19
K02600	N utilization substance protein A	*nusA*	[5]Transcription	0.1126	0.0726	1.02	[1]
K02518	Translation initiation factor IF-1	*infA*	[5]Translation, ribosomal structure, and biogenesis	0.0346	0.0646	0.98	21

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 65.4% of the variation in ADG (average daily gain). Rows colored in gray correspond to genes simultaneously identified for both FCR (feed conversion ratio) and ADG prediction.

Metabolic Pathways of Microbial Genes Associated With Efficiency Traits

Our results indicate that most proteins encoded by microbial genes identified for the prediction of FCR were generally involved in carbohydrates metabolism and transport. For example, *aguA* and K01188 are involved in biomass conversion, through the degradation of hemicelluloses and lignocelluloses and lactate biosynthesis (Cairns and Esen, 2010; Lee et al., 2012; Michlmayr and Kneifel, 2014; Li, 2015). Microbial genes *xylE*, *aguA*, and *uidA* are involved in xylan degradation, the main component of hemicellulose (Lee et al., 2012; Fliegerova et al., 2015). Xylose needs to be taken up by a transporter (putatively associated with *xylE*) before it is metabolized, and it has been recognized as a rate-controlling step in bacterial metabolism (Chaillou and Pouwels, 1999). Furthermore, microbial genes such as *uidA* [previously identified by Roehe et al. (2016)], directly involved in carbohydrate metabolism pathways like pentose and glucuronate interconversions and galactose metabolism, are coupled with NAD or NADP oxidoreduction, important for regulating the flux of carbon and energy sources in microorganisms (Spaans et al., 2015). In addition, *punA* (i.e., purine-nucleoside phosphorylase) is involved in the metabolism of nucleotides, nicotinate and nicotinamide (vitamin B3), which also contain NAD and NADP, and is therefore important in carbohydrate, protein, and lipid

TABLE 4 | Summary of microbial genes identified for the prediction of RFI.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K03406	Methyl-accepting chemotaxis protein	*mcp*	[2]Two-component system; bacterial chemotaxis	0.0225	–0.0510	1.26	[1]
K03413	Two-component system, chemotaxis family, response regulator CheY	*cheY*	[2]Two-component system; bacterial chemotaxis	0.0018	0.0478	1.16	[1]
K01534	Cd2+/Zn2+-exporting ATPase	*zntA*	[5]Cation-transporting ATPase activity; metal ion binding; nucleotide binding	0.0211	–0.0653	1.16	04
K07258	serine-type D-Ala-D-Ala carboxypeptidase (penicillin-binding protein 5/6)	*dacC*	[2]Metabolic pathways; Peptidoglycan biosynthesis	0.0049	–0.0375	1.14	[1]
K07301	Cation:H+ antiporter	*yrbG*	[3]Inorganic ion transport and metabolism	0.0096	–0.0145	1.09	04
K04720	Threonine-phosphate decarboxylase	*cobD*	[2]Porphyrin and chlorophyll metabolism	0.0034	–0.0501	1.06	04
K03407	Two-component system, chemotaxis family, sensor kinase CheA	*cheA*	[2]Two-component system; bacterial chemotaxis	0.0048	–0.0236	1.04	[1]
K00595	Precorrin-6Y C5,15-methyltransferase (decarboxylating)	*cobL*	[2]Metabolic pathways; porphyrin and chlorophyll metabolism	0.0078	0.0223	1.02	04
K01571	Oxaloacetate decarboxylase, alpha subunit	*oadA*	[2]Metabolic pathways; pyruvate metabolism	0.0165	–0.0501	0.96	04
K02057	Simple sugar transport system permease protein	*ABC.SS.P*	[3]Carbohydrate transport and metabolism	0.0023	–0.1375	0.96	20
K02390	Flagellar hook protein FlgE	*flgE*	[2]Flagellar assembly	0.0015	–0.0376	0.87	[1]
K02417	Flagellar motor switch protein FliN/FliY	*fliN*	[2]Bacterial chemotaxis; flagellar assembly	0.0018	–0.1120	0.77	[1]
K03738	Aldehyde:ferredoxin oxidoreductase	*aor*	[2]Metabolic pathways; Microbial metabolism in diverse environments; Pentose phosphate pathway; Carbon metabolism	0.0144	–0.0657	0.68	NC
K02009	Cobalt transport protein	*cbiN*	[2]ABC transporters	0.0074	–0.1126	0.67	01
K01709	CDP-glucose 4,6-dehydratase	*rfbG*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism	0.0041	0.2549	1.46	[1]
K00978	Glucose-1-phosphate cytidylyltransferase	*rfbF*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism; starch and sucrose metabolism	0.0042	0.2056	1.23	[1]
K14113	Energy-converting hydrogenase B subunit D	*ehbD*	–	0.0010	0.1703	1.00	NC

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. [1]Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: [2]KEGG database, [3]NCBI database, [4]BioCyc database, and [5]UniProt database. The genes in this table explained 65.6% of the variation in RFI (residual feed intake). Rows colored in grey correspond to genes simultaneously identified for both RFI and DFI (daily feed intake) prediction.

metabolism reactions. Positive effects of vitamin B3 have been previously observed in healthy rumen microbiomes in beef and dairy cattle (Aschemann et al., 2012; Luo et al., 2017). Microbial genes *uidA* and *punA* were more abundant in efficient animals.

Proteins encoded by *lctP*, K01188, and *ptb*, involved in lactate transport and cellulose and butyrate metabolism, respectively, could be involved in host–microbiome crosstalk mechanisms in cattle due to their participation in metabolic pathways that involve the release of H^+, such as lactate metabolism, potentially reducing microbial fiber-degrading activity and consequently slowing digestion and rumen emptying rate, causing a decrease in appetite (Moran, 2005b). Furthermore, beta-glucosidase is widely present in lactic acid bacteria and is thought to interact with the human host (Michlmayr and Kneifel, 2014). Butyrate has been shown in rats to directly activate the intestinal gluconeogenesis genes in enterocytes *via* an increase in cationic antimicrobial peptides (cAMP, De Vadder et al., 2014). In contrast, *glo1* (more abundant in FCR-H) is involved in methylglyoxal degradation, which is a highly toxic substance that decreases bacterial cell viability, and is produced by bacteria when there is carbohydrate excess and nitrogen limitation (Russell, 1993). Therefore, *glo1* is a strong candidate biomarker of rumen microbiome difference in less efficient animals (i.e., FCR-H).

The microbial gene with highest impact in prediction of ADG was *amiABC*, which is mainly involved in the peptidoglycan turnover through cleavage of glyosidic bonds and release of amino acids and cAMP resistance (Uehara and Park, 2008; Uehara et al., 2010). Some bacteria (mostly pathogenic) have evolved mechanisms of resistance, such as decreased affinity to cAMPs (Anaya-López et al., 2013), and the higher abundance of *amiABC* in animals with lower ADG may be indicative of higher abundance of pathogens, which can cause inflammatory response in the rumen potentially reducing nutrient use and absorption (Reynolds et al., 2017). Brown et al. (2003) demonstrated that acetate and propionate are agonists of the human receptors GPR43 and GPR41, and Hong et al. (2005) proposed that acetate

TABLE 5 | Summary of microbial genes identified for the prediction of DFI.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K00370	Nitrate reductase 1, alpha subunit	*narG*	[2]Microbial metabolism in diverse environments; nitrogen metabolism; two-component system	0.0022	–0.2272	1.22	[1]
K01858	Myo-inositol-1-phosphate synthase	*INO1*	[2]Metabolic pathways; biosynthesis of antibiotics; streptomycin biosynthesis; inositol phosphate metabolism	0.0542	–0.0459	1.14	[1]
K03685	Ribonuclease III	*rnc*	[2]Ribosome biogenesis in eukaryotes; proteoglycans in cancer	0.0288	–0.0097	1.13	[1]
K00613	Glycine amidinotransferase	*GATM*	[2]Metabolic pathways; glycine, serine and threonine metabolism; arginine and proline metabolism	0.0019	–0.1417	1.09	[1]
K02428	XTP/dITP diphosphohydrolase	*rdgB*	[2]Metabolic pathways; purine metabolism	0.0147	–0.0216	0.94	02
K03602	Exodeoxyribonuclease VII small subunit	*xseB*	[2]Mismatch repair	0.0035	0.0803	0.94	02
K03210	Preprotein translocase subunit YajC	*yajC*	[2]Bacterial secretion system; quorum sensing; protein export	0.0069	0.1317	0.93	[1]
K12340	Outer membrane channel protein TolC	*tolC*	[2]Beta-lactam resistance; cationic antimicrobial peptide (CAMP) resistance; two-component system; bacterial secretion system; plant–pathogen interaction; pertussis	0.0157	0.0068	0.92	02
K03043	DNA-directed RNA polymerase subunit beta	*rpoB*	[2]Metabolic pathways; purine metabolism; pyrimidine metabolism; RNA polymerase	1.2470	–0.0995	0.91	NC
K04751	Nitrogen regulatory protein P-II 1	*glnB*	[2]Two-component system	0.0151	0.0613	0.91	02
K03625	N utilization substance protein B	*nusB*	[3]Transcription termination	0.0135	0.0766	0.91	02
K06178	Ribosomal large subunit pseudouridine synthase B	*rluB*	[3]Translation, ribosomal structure, and biogenesis	0.0693	–0.0038	0.85	02
K05349	Beta-glucosidase	*bglX*	[2]Metabolic pathways; biosynthesis of secondary metabolites; cyanoamino acid metabolism; starch and sucrose metabolism; phenylpropanoid biosynthesis	0.2272	0.0063	0.84	[1]
K05515	Penicillin-binding protein 2	*mrdA*	[2]Peptidoglycan biosynthesis; beta-lactam resistance	0.0295	0.0214	0.82	02
K04764	Integration host factor subunit alpha	*ihfA*	[3]DNA binding: replication, recombination, and repair	0.0041	0.0306	0.80	02
K01709	CDP-glucose 4,6-dehydratase	*rfbG*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism	0.0041	0.2412	1.53	[1]
K00978	Glucose-1-phosphate cytidylyltransferase	*rfbF*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism; starch and sucrose metabolism	0.0042	0.2634	1.43	[1]
K14113	Energy-converting hydrogenase B subunit D	*ehbD*	–	0.0010	0.1594	1.16	NC

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 72.9% of the variation in DFI (daily feed intake). Rows colored in gray correspond to genes simultaneously identified for both RFI (residual feed intake) and DFI prediction.

and propionate induce lipid accumulation and inhibition of lipolysis through the GPR43 receptor in mice. These genes are also part of the bovine genome, where they mediate an inhibitory effect of acetate, propionate, and butyrate on cAMP signaling (Wang et al., 2009). This could indicate that, in less efficient animals (lower ADG), the lower amount of acetate, propionate, and butyrate may lead to decreased inhibition of lipolysis by the host, which potentially results in lower ADG. Alternatively, the lower amount of VFAs in these animals may lead to decreased inhibition of cAMP signaling and increased release of cAMPs by the host to the rumen. The cAMPs act primarily on organisms without effective resistance mechanisms, consequently increasing the relative abundance of cAMP-resisting organisms and of the microbial genes encoding for the resistance. Two other microbial genes identified in the present research are part of the cAMP resistance pathway—*lpxA* and *tolC* (associated with FCR and DFI, respectively). Although all three genes (*amiABC*, *lpxA*, and *tolC*) are part of the same pathway, they present opposite tendencies—while *lpxA* and *tolC* are proportionally highly abundant in animals with higher ADG and lower FCR, *amiABC* is relatively highly abundant in animals with lower ADG and higher FCR. The gene *lpxA* is related to lipid A integration in the cell wall, as a preventive measure against the hosts' immune system, and *tolC* is involved in the efflux of antibiotics (Raetz et al., 2007; Zgurskaya et al., 2011). This could be indicative of the different cAMP resistance mechanisms evolved by bacterial

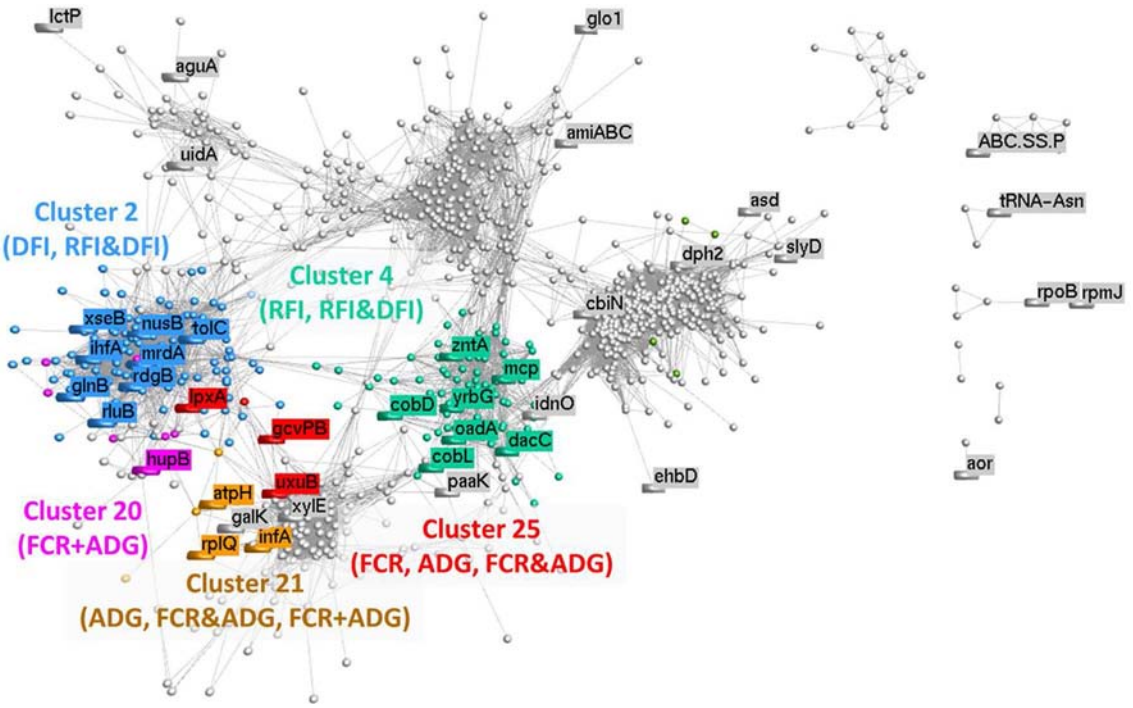


FIGURE 5 | Correlation network analysis of metagenomic data: Each node represents a vector of relative abundances of each microbial gene in all 42 animals, and the edges represent a correlation between the microbial genes. A minimum correlation threshold of 0.80 was applied to the network. Different colors illustrate different clusters, which were calculated using MCL method (inflation: 2; preinflation: 2; scheme: 6). Clusters identified by numbers were found to be significantly ($P < 0.05$) enriched for microbial genes identified for the traits whose abbreviations are between brackets (FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; FCR&ADG, set including microbial genes identified for prediction of either FCR and/or ADG; RFI&DFI, set including microbial genes identified for prediction of RFI and/or DFI; FCR+ADG, set including microbial genes simultaneously identified for prediction of both traits FCR and ADG).

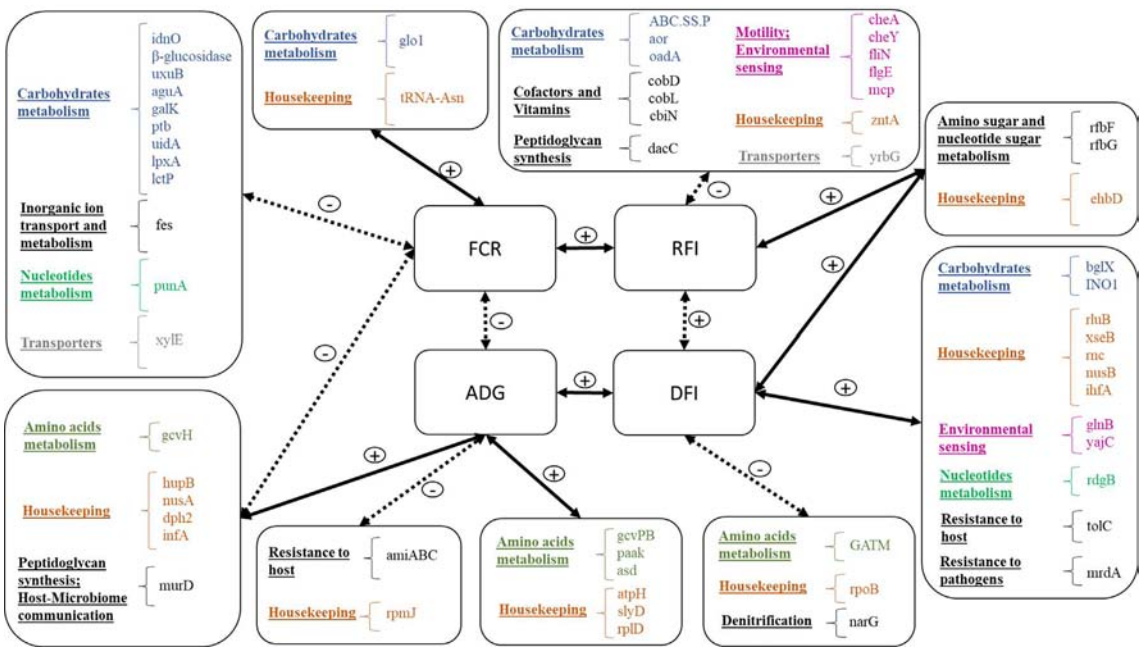


FIGURE 6 | Summary of microbial genes identified for the prediction of each trait: Traits are located in the four central boxes: FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake. Solid lines represent positive correlations, and dotted lines represent negative correlations. Microbial genes are listed in the outside boxes, organized by general function, and each general function is represented by a different color.

organisms, which include modification of the cell external surface, efflux pumps, and biosynthesis and crosslinking of cell envelope components (Nizet, 2006).

The set of microbial genes associated with ADG included mostly housekeeping genes and genes related to amino acid metabolism and transport. Artegoitia et al. (2017) found a link between ruminal aromatic amino acids synthesis such as phenylalanine and high ADG in beef steers. For example, *paak* [previously mentioned by Kamke et al. (2016) related to sheep with high production of methane] and *asd* encode proteins that respectively catalyze phenylalanine and phenylacetate (related to aspartate degradation and biosynthesis of amino acids including threonine), with release of H^+. In the current research, both of these genes were positively correlated to ADG, which is supported by the positive correlations between ADG and dry matter intake (DMI), between DMI and methane emissions, and between methane emissions and body weight measurements (weaning weight, yearling weight, and final weight), previously observed in cattle (Koots et al., 1994; Arthur et al., 2001; Herd et al., 2014).

Some housekeeping genes were simultaneously identified for the prediction of FCR and ADG, such as protein translation from diphthamide (*dph2*) or peptidoglycan biosynthesis (*murD*), both more abundant in efficient animals (higher ADG and lower FCR). The importance of diphthamide biosynthesis in archaea is not yet fully known (Narrowe et al., 2018). Microbial gene *murD* is related to the glutamate–glutamine cycle, an important appetite regulator in humans (Delgado, 2013), but in the present research, it was not associated to DFI.

Proteins encoded by microbial genes associated with RFI are mostly related to chemotaxis (*cheA* and *cheY*), detoxification (Cd2+/Zn2+-exporting ATPase, *zntA*), and vitamin B12 production (*cbiN*, *cobD*, and *cobL*). The negative correlation of microbial genes involved in chemotaxis and motility with RFI may suggest an increased microbial metabolism in efficient animals, derived from their ability to sense chemical gradients in their surrounding environment and to react accordingly, i.e., moving closer to nutrients (Rajagopala et al., 2007). Microbial gene *zntA* was also more abundant in efficient animals and plays a role in the homeostasis of transition metals (Cd^{2+}, Zn^{2+}), participating in functional pathways ranging from cellular respiration to gene expression (Fraústro da Silva and Williams, 2001). Finally, higher relative abundance of microbial genes involved in vitamin B12 production (*cbiN*, *cobD*, and *cobL*) was observed in more efficient animals. This essential cofactor needs to be taken up directly from the diet or to be made available for animal absorption by the rumen microbial organisms because it is not produced by eukaryotes (Warren et al., 2002). Furthermore, vitamin B12 has been previously associated with increased cobalt content on high-fiber diets and increased VFA, such as acetate (Beaudet et al., 2017), which may affect the animals' appetite (Frost et al., 2014), in line with our observation of higher relative abundance of these genes in more efficient animals, i.e., animals with lower feed intake than expected.

The four most important microbial genes identified for the prediction of DFI included the three microbial genes also identified for prediction of RFI (*rfbG*, *rfbF*, and *ehbD*) and *narG*. Microbial genes *rfbG* and *rfbF* (VIP > 1.4) are part of the *rfc* region (Morona et al., 1994) and are related to nucleotide sugar metabolism, which is necessary for the production of microbial lipopolysaccharide (LPS). LPS is a major virulence factor of Gram-negative bacteria, particularly due to the O-antigen, paramount for host colonization and niche adaptation by bacterial organisms, due to its part in the protection from host immune response (Reeves, 1995; Samuel and Reeves, 2003; Geue et al., 2017). Both genes *rfbG* and *rfbF* showed a positive correlation to RFI and DFI, supporting our hypothesis that the use of energy to stimulate the innate immune system against pathogens increases DFI and reduces feed conversion efficiency as determined by RFI (Neal et al., 1991; Jing et al., 2014; Vigors et al., 2016). Other microbial genes positively correlated to DFI were found to be involved in resistance mechanisms, such as the penicillin-binding protein 2-encoding gene (*mrdA*), which belongs to the peptidoglycan and beta-lactam resistance metabolic pathways. These proteins are transpeptidases or carbopeptidases involved in peptidoglycan metabolism and have an important role against beta-lactam resistance (Zapun et al., 2008). The microbial gene myo-inositol-1-phosphate synthase (*INO1*) is related to antibiotic biosynthesis, including streptomycin. Microbial gene *ehbD* is a subunit of the energy-converting hydrogenase B, found in methanogens such as *Methanococcus maripaludis*. This microbial gene is important due to its role in autotrophic CO_2 assimilation (Porat et al., 2006), having implications for microbial growth. Furthermore, *narG*, part of the *narGHIJ* operon, essential for some microorganisms to gather energy under anaerobic conditions by the reduction in nitrate to nitrite in a denitrification process (Blasco et al., 1990; Latham et al., 2016), was proportionally more abundant in animals with low DFI.

The microbial gene *nusB* (associated with DFI) is part of a set of *nus* genes, which also includes *nusA* (identified for prediction of FCR and ADG). Genes in the *nus* complex are involved in transcription termination and antitermination processes, such as Rho-dependent transcriptional termination (Torres et al., 2004), which is the regulatory mechanism involved in the efficient transcription of the tryptophan operon (Farnham et al., 1982; Kuroki et al., 1982; Prasch et al., 2009). The *nus*-complex microbial genes were found to be relatively more abundant in efficient animals. This association may be due to the influence of the *nus* genes, which extends from the ribosomal operons to the tryptophan operon and constitutes a good example of how termination and antitermination processes can control gene expression, occurring during RNA transcription, and potentially positively impacting bacterial growth and rumen fermentation processes.

Although microbial genes *amiABC*, *tolC*, *glo1, rfbF*, *rfbG*, *and lpxA* were identified in the present research for the prediction of different traits, all are associated with bacterial defense mechanisms either from other bacteria or from the host. The majority of these genes had higher abundance in less efficient animals. This suggests that the presence of either bacterial pathogens in the rumen or antibiotics produced as host immune responses might represent a significant energy sink, impairing feed conversion efficiency.

Further improvement of prediction of feed conversion traits using metagenomic information may be achieved through

the integration of protein, enzyme, and pathway data from the Hungate collection (Seshadri et al., 2018) and the large rumen metagenomic reference dataset (Stewart et al., 2018).

CONCLUSIONS

The results presented here suggest that relative abundances of rumen microbial genes may be highly informative predictors of feed conversion efficiency, growth rate, and feed intake, which are labor intensive, time consuming, and expensive traits to record. Most microbial genes identified for the prediction of traits in this research were trait specific. Microbial genes related to cellulose and hemicellulose degradation, vitamin B12 synthesis, and amino acids metabolism were associated to enhanced feed conversion efficiency (FCR or RFI), while those involved in nucleotide sugars metabolism, pathogen LPS synthesis, cAMP resistance, and degradation of toxic compounds were associated with inefficient feed conversion. Furthermore, we identified specific microbial genes encoding proteins related to the crosstalk between the microbiome and the host cells, such as *murD* and *amiABC*, and associated to gene expression regulatory mechanisms, such as *nusA* and *nusB*. Thus, our results provide a deeper understanding of the potential influence of the rumen microbiome on the feed conversion efficiency of its host, highlighting specific enzymes involved in metabolic pathways that reflect the complex functional networks impacting the conversion of feed into animal products such as meat.

AUTHOR CONTRIBUTIONS

RR and MW conceived and designed the overall study, and JL, MA, TF, and RR conceived, designed, and executed the bioinformatics analysis. RS and MW carried out the bioinformatics to obtain the rumen microbial gene abundances. C-AD, TS, RD, and AW provided essential insight into feed conversion efficiency, rumen metabolism, nutrition, and microbiology. JL, MA, and RR wrote the initial draft, and subsequently, all authors contributed intellectually to the interpretation and presentation of the results.

FUNDING

The project was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC BB/N01720X/1 and BB/N016742/1) and by the Scottish Government (RESAS Division) as part of the 2016–2021 commission. The research is based on data from experiments funded by the Scottish Government as part of the 2011–2016 commission, Agriculture and Horticulture Development Board (AHDB) Beef & Lamb, Quality Meat Scotland (QMS), and Department for Environment Food & Rural Affairs (Defra).

ACKNOWLEDGMENTS

We thank Dr Irene Cabeza Luna, Laura Nicoll, Lesley Deans, and Claire Broadbent for the excellent technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00701/full#supplementary-material

REFERENCES

Anaya-López, J. L., López-Meza, J. E., and Ochoa-Zarzosa, A. (2013). Bacterial resistance to cationic antimicrobial peptides. *Crit. Rev. Microbiol.* 39, 180–195. doi: 10.3109/1040841X.2012.699025

Arora, T., Sharma, R., and Frost, G. (2011). Propionate. Anti-obesity and satiety enhancing factor? *Appetite* 56, 511–515. doi: 10.1016/j.appet.2011.01.016

Artegoitia, V. M., Foote, A. P., Lewis, R. M., and Freetly, H. C. (2017). Rumen fluid metabolomics analysis associated with feed efficiency on crossbred steers. *Sci. Rep.* 7, 1–14. doi: 10.1038/s41598-017-02856-0

Arthur, J. P. F., and Herd, R. M. (2008). Residual feed intake in beef cattle. *Rev. Bras. Zootec.* 37, 269–279. doi: 10.1590/S1516-35982008001300031

Arthur, P. F., Renand, G., and Krauss, D. (2001). Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. *Livest. Prod. Sci.* 68, 131–139. doi: 10.1016/S0301-6226(00)00243-8

Aschemann, M., Lebzien, P., Hüther, L., Südekum, K.-H., and Dänicke, S. (2012). Effect of niacin supplementation on rumen fermentation characteristics and nutrient flow at the duodenum in lactating dairy cows fed a diet with a negative rumen nitrogen balance. *Arch. xAnim. Nutr.* 66, 303–318. doi: 10.1080/1745039x.2012.697353

Auffret, M. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., John Wallace, R., Freeman, T. C., et al. (2017). The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle. *Microbiome* 5, 159–169. doi: 10.1186/s40168-017-0378-z

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Beaudet, V., Gervais, R., Graulet, B., Nozière, P., Doreau, M., Fanchone, A., et al. (2017). Effects of dietary nitrogen levels and carbohydrate sources on apparent ruminal synthesis of some B vitamins in dairy cows. *J. Dairy Sci.* 99, 2730–2739. doi: 10.3168/jds.2017-100-1-0849

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl. Acad. Sci.* 107, 18933–18938. doi: 10.1073/pnas.1007028107

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590. doi: 10.1152/physrev.1990.70.2.567

Berry, D. P., and Crowley, J. J. (2013). Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91, 1594–1613. doi: 10.2527/jas.2012-5862

Blasco, F., Iobbi, C., Ratouchniak, J., Bonnefoy, V., and Chippaux, M. (1990). Nitrate reductases of *Escherichia coli*: sequence of the second nitrate reductase and comparison with that encoded by the narGHJI operon. *MGG Mol. Gen. Genet.* 222, 104–111.

Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., et al. (2003). The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* 278, 11312–11319. doi: 10.1074/jbc.m211609200

Brulc, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci.* 106, 1948–1953. doi: 10.1073/pnas.0806191105

Cairns, J. R. K., and Esen, A. (2010). β-Glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2

Chaillou, P., and Pouwels, P. H. (1999). Transport of D-xylose in *Lactobacillus pentosus*, *Lactobacillus casei*, and *Lactobacillus plantarum*: evidence for a mechanism of facilitated diffusion *via* the phosphoenolpyruvate: mannose phosphotransferase system. *Society* 181, 4768–4773.

Cherbut, C. (2003). Motor effects of short-chain fatty acids and lactate in the gastrointestinal tract. *Proc. Nutr. Soc.* 62, 95–99. doi: 10.1079/PNS2002213

Creevey, C. J., Kelly, W. J., Henderson, G., and Leahy, S. C. (2014). Determining the culturability of the rumen bacterial microbiome. *Microb. Biotechnol.* 7, 467–479. doi: 10.1111/1751-7915.12141

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., et al. (2014). Microbiota-generated metabolites promote metabolic benefits *via* gut–brain neural circuits. *Cell* 156, 84–96. doi: 10.1016/j.cell.2013.12.016

Delgado, T. C. (2013). Glutamate and GABA in appetite regulation. *Front. Endocrinol.* 4, 103. doi: 10.3389/fendo.2013.00103

Duthie, C.-A., Rooke, J. A., Troy, S., Hyslop, J. J., Ross, D. W., Waterhouse, A., et al. (2015). Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. *Animal* 10, 786–795. doi: 10.1017/S1751731115002657

Duthie, C. A., Troy, S. M., Hyslop, J. J., Ross, D. W., Roehe, R., and Rooke, J. A. (2017). The effect of dietary addition of nitrate or increase in lipid concentrations, alone or in combination, on performance and methane emissions of beef cattle. *Animal* 12, 280–287. doi: 10.1017/S175173111700146X

Eurostat. (2018). Eurostat online table Bovine, alive, EU-28. Luxembourg: Eurostat. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apro_mt_lscatl&lang=en (Accessed October 16, 2018).

Farnham, P. J., Greenblatt, J., and Platt, T. (1982). Effects of NusA protein on transcription termination in the tryptophan operon of *Escherichia coli. Cell* 29, 945–951. doi: 10.1016/0092-8674(82)90457-3

Fliegerova, K., Kaerger, K., Kirk, P., and Voigt, K. (2015). "7. Rumen fungi," in *Rumen Microbiology: From Evolution to Revolution* (India: Springer India), 97–112. doi: 10.1007/978-81-322-2401-3

Fraústro da Silva, J., and Williams, R. (2001). *The biological chemistry of the elements: the inorganic chemistry of life*. New York: Oxford University Press.

Freeman, T. C., Goldovsky, L., Brosch, M., van Dongen, S., Maziere, P., Grocock, R. J., et al. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput. Biol.* 3, e206 (11 pages). doi: 10.1371/journal.pcbi.0030206

Frost, G., Sleeth, M. L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., et al. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* 5, 3611. doi: 10.1038/ncomms4611

Gerber, P. J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., et al. (2013). Tackling climate change through livestock - A global assessment of emissions and mitigation opportunities. Rome: Food and Agriculture Organization of the United Nations (FAO).

Geue, L., Menge, C., Eichhorn, I., Semmler, T., Wieler, L. H., Pickard, D., et al. (2017). Evidence for contemporary switching of the O-antigen gene cluster between shiga toxin-producing escherichia coli strains colonizing cattle. *Front. Microbiol.* 8, 424. doi: 10.3389/fmicb.2017.00424

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F. et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Guan, L. L., Nkrumah, J. D., Basarab, J. A., and Moore, S. S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol. Lett.* 288, 85–91. doi: 10.1111/j.1574-6968.2008.01343.x

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Global Rumen Census Collaborators, G., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5, 14567. doi: 10.1038/srep14567

Herd, R. M., Arthur, P. F., Donoghue, K. A., Bird, S. H., Bird-Gardiner, T., and Hegarty, R. S. (2014). Measures of methane production and their phenotypic relationships with dry matter intake, growth, and body composition traits in beef cattle. *J. Anim. Sci.* 92, 5267–5274. doi: 10.2527/jas.2014-8273

Hernandez-Sanabria, E., Goonewardene, L. A., Wang, Z., and Durunna, O. N. (2011). Impact of feed efficiency and diet on the adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl. Environ. Microbiol.* 1, 1203–1214. doi: 10.1128/AEM.05114-11

Hong, Y.-H., Nishimura, Y., Hishikawa, D., Tsuzuki, H., Miyahara, H., Gotoh, C., et al. (2005). Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology* 146, 5092–5099. doi: 10.1210/en.2005-0545

Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* 7, e33306. doi: 10.1371/journal.pone.0033306

Jami, E., White, B. A., and Mizrahi, I. (2014). Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* 9, e85423. doi: 10.1371/journal.pone.0085423

Jiang, X., Neal, B., Santiago, F., Lee, S. J., and Reeves, P. R. (2006). Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella serovar typhimuriunfi* (strain LT2). *Mol. Microbiol.* 5, 695–713. doi: 10.1111/j.1365-2958.1991.tb00741.x

Jing, L., Zhang, R., Liu, Y., Zhu, W., and Mao, S. (2014). Intravenous lipopolysaccharide challenge alters ruminal bacterial microbiota and disrupts ruminal metabolism in dairy cattle. *Br. J. Nutr.* 112, 170–182. doi: 10.1017/S000711451400066X

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4, 56. doi: 10.1186/s40168-016-0201-2

Kanehisa Laboratories (2018). KEGG MAPPER. Available at: https://www.genome.jp/kegg/tool/map_pathway1.html [Accessed October 1, 2018]

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1016/j.meegid.2016.07.022

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., et al. (2017). The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 201, 1–9. doi: 10.1093/bib/bbx085

Koots, K. R., Gibson, J. P., and Wilton, J. W. (1994). Analyses of published genetic parameter estimates for beef production traits. 2. Phenotypic and genetic correlations. *Anim. Breed. Abstr.* 62, 825–853. Available at: http://agris.fao.org/agris-search/search.do?recordID=GB9508564

Kuroki, K., Ishii, S., Kano, Y., Miyashita, T., Nishi, K., and Imamoto, F. (1982). Involvement of the nusA and nusB gene products in transcription of *Escherichia coli* tryptophan operon *in vitro. MGG Mol. Gen. Genet.* 185, 369–371. doi: 10.1007/BF00330816

Latham, E. A., Anderson, R. C., Pinchak, W. E., and Nisbet, D. J. (2016). Insights on alterations to the rumen ecosystem by nitrate and nitrocompounds. *Front. Microbiol.* 7, 228. doi: 10.3389/fmicb.2016.00228

Lee, C. C., Kibblewhite, R. E., Wagschal, K., Li, R., and Orts, W. J. (2012). Isolation of α-glucuronidase enzyme from a rumen metagenomic library. *Protein J.* 31, 206–211. doi: 10.1007/s10930-012-9391-z

Li, M., Penner, G. B., Hernandez-Sanabria, E., Oba, M., and Guan, L. L. (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *J. Appl. Microbiol.* 107, 1924–1934. doi: 10.1111/j.1365-2672.2009.04376.x

Li, R. W. (2015). "16. Rumen metagenomics," in *Rumen Microbiology: From Evolution to Revolution* (India: Springer India), 223–245. doi: 10.1007/978-81-322-2401-3

Luo, D., Gao, Y., Lu, Y., Qu, M., Xiong, X., Xu, L., et al. (2017). Niacin alters the ruminal microbial composition of cattle under high-concentrate condition. *Anim. Nutr.* 3, 180–185. doi: 10.1016/j.aninu.2017.04.005

Lyons, T., Bielak, A., Doyle, E., and Kuhla, B. (2018). Variations in methane yield and microbial community profiles in the rumen of dairy cows as they pass through stages of first lactation. *J. Dairy Sci.* 101, 5102–5114. doi: 10.3168/jds.2017-14200

Maldini, G., and Allen, M. S. (2018). Temporal effects of ruminal propionic acid infusion on feeding behavior of Holstein cows in the postpartum period. *J. Dairy Sci.* 101, 3077–3084 doi: 10.3168/jds.2017-13857

Meale, S. J., Auffret, M. D., Watson, M., Morgavi, D. P., Cantalapiedra-Hijar, G., Duthie, C. A., et al. (2018). Fat accretion measurements strengthen the relationship between feed conversion efficiency and nitrogen isotopic discrimination while rumen microbial genes contribute little. *Sci. Rep.* 8, 3854. doi: 10.1038/s41598-018-22103-4

Michlmayr, H., and Kneifel, W. (2014). β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. *FEMS Microbiol. Lett.* 352, 1–10. doi: 10.1111/1574-6968.12348

Moran, J. (2005a). "Chapter 5: How the rumen works," in *Tropical dairy farming: feeding management for small holder dairy farmers in the humid tropics*. (Collingwood: Csiro Publishing), 41–49. doi: 10.1016/0022-1694(84)90033-7

Moran, J. (2005b). "Chapter 13. Problems with unbalanced diets," in *Tropical Dairy Farming: Feeding Management for Small Holder Dairy Farmers in the Humid Tropics.* doi: 10.1071/9780643093133

Morona, R., Mavris, M., Fallarino, A., and Manning, P. A. (1994). Characterization of the rfc region of *Shigella flexneri*. *J. Bacteriol.* 176, 733–747. doi: 10.1128/jb.176.3.733-747.1994

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A., and Freetly, H. C. (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10, e0129174. doi: 10.1371/journal.pone.0129174

Narrowe, A. B., Spang, A., Stairs, C. W., Caceres, E. F., Baker, B. J., Miller, C. S., et al. (2018). Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in archaea and parabasalids. *Genome Biol. Evol.* 10, 2380–2393. doi: 10.1093/gbe/evy154

Neal, B., Jiang, X.-M., Lee, S. J., Santiago, F., Reeves, P. R., and Romana, L. K. (1991). Structure and sequence of the rfb (O antigen) gene cluster of Salmonella serovar typhimurium (strain LT2). *Mol. Microbiol.* 5, 695–713. doi: 10.1111/j.1365-2958.1991.tb00741.x

Nielsen, M. K., MacNeil, M. D., Dekkers, J. C. M., Crews, D. H., Jr., Rathje, T. A., Enns, R. M., et al. (2013). Review: Life-cycle, total-industry genetic improvement of feed efficiency in beef cattle: Blueprint for the Beef Improvement Federation. *Prof. Anim*. Sci. 29, 559–565. doi: 10.15232/s1080-7446(15)30285-0

Nizet, V. (2006). Antimicrobial peptide resistance mechanisms of human bacterial pathogens. *Curr Issues Mol Biol* 8, 11–26 Available at: https://www.ncbi.nlm.nih.gov/pubmed/16450883.

Oba, M., and Allen, M. S. (2003). Intraruminal infusion of propionate alters feeding behavior and decreases energy intake of lactating dairy cows. *J. Nutr* 133, 1094–1099. doi: 10.3168/jds.S0022-0302(03)73889-2

Porat, I., Kim, W., Hendrickson, E. L., Xia, Q., Zhang, Y., Wang, T., et al. (2006). Disruption of the operon encoding Ehb hydrogenase limits anabolic CO 2 assimilation in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 188, 1373–1380. doi: 10.1128/JB.188.4.1373-1380.2006

Prasch, S., Jurk, M., Washburn, R. S., Gottesman, M. E., Wöhrl, B. M., and Rösch, P. (2009). RNA-binding specificity of *E. coli* NusA. *Nucleic Acids Res*. 37, 4736–4742. doi: 10.1093/nar/gkp452

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, S., Manichanh, C., et al., (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Raetz, C. R. H., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007). Lipid A modification systems in Gram-negative bacteria. *NIH Public Access* 76, 295–329. doi: 10.1146/annurev.biochem.76.010307.145803

Rajagopala, S. V., Titz, B., Goll, J., Parrish, J. R., Wohlbold, K., McKevitt, M. T., et al. (2007). The protein network of bacterial motility. *Mol. Syst. Biol.* 3, 128. doi: 10.1038/msb4100166

Reeves, P. (1995). Role of O-antigen variation in the immune response. *Trends Microbiol*. 3, 381–386. doi: 10.1016/S0966-842X(00)88983-0

Reynolds, J. G., Foote, A. P., Freetly, H. C., Oliver, W. T., and Lindholm-Perry, A. K. (2017). Relationships between inflammation- and immunity-related transcript abundance in the rumen and jejunum of beef steers with divergent average daily gain. *Anim. Genet*. 48, 447–449. doi: 10.1111/age.12546

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi: 10.1371/journal.pgen.1005846

Rooke, J. A., Wallace, R. J., Duthie, C.-A., McKain, N., Souza, S. M. de, Hyslop, J. J., et al. (2014). Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *Br. J. Nutr*. 112, 398–407. doi: 10.1017/s0007114514000932

Ross, E. M., Moate, P. J., Marett, L. C., Cocks, B. G., and Hayes, B. J. (2013). Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* 8, e73056. doi: 10.1371/journal.pone.0073056

Russell, J. B. (1993). Glucose toxicity in *Prevotella ruminicola*: methylglyoxal accumulation and its effect on membrane physiology. *Appl. Environ. Microbiol.* 59, 2844–2850. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8215358

Russell, J. B. (2001). Factors that alter rumen microbial ecology. *Science* 292, 1119–1122. doi: 10.1126/science.1058830

Russell, J. B., and Hespell, R. B. (1981). Microbial rumen fermentation. *J. Dairy Sci*. 64, 1153–1169. doi: 10.3168/jds.S0022-0302(81)82694-X

Sakata, T., and Tamate, H. (1979). Rumen epithelial cell proliferation accelerated by propionate and acetate. *J. Dairy Sci.* 61, 1109–1113. doi: 10.3168/jds.S0022-0302(78)83694-7

Samuel, G., and Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 338, 2503–2519. doi: 10.1016/j.carres.2003.07.009

Sasson, G., Ben-Shabat, S., Seroussi, E., Doron-Faigenboim, A., Shterzer, N., Yaacoby, S., et al. (2017). Heritable bovine rumen bacteria are phylogenetically related and correlated with the cow's capacity to harvest. *Am. Soc. Microbiol*. 8, e00703–17. doi: 10.1128/mBio.00703-17

Scollan, N. D., Morrisson, S., Popova, M., Attwood, G. T., Creevey, C. J., Tapio, I., et al. (2018). Addressing global ruminant agricultural challenges through understanding the rumen microbiome: past, present, and future. *Front. Microbiol.* 9, 2161. doi: 10.3389/fmicb.2018.02161

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 36, 359–367. doi: 10.1038/nbt.4110

Shabat, S., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M. E., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* 10, 2958–2972. doi: 10.1038/ismej.2016.62

Spaans, S. K., Weusthuis, R. A., van der Oost, J., and Kengen, S. W. M. (2015). NADPH-generating systems in bacteria and archaea. *Front. Microbiol.* 6, 742. doi: 10.3389/fmicb.2015.00742

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun*. 9, 870. doi:10.1038/s41467-018-03317-6

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584. doi: 10.1111/j.1462-2920.2009.01982.x

Torres, M., Balada, J. M., Zellars, M., Squires, C., and Squires, C. L. (2004). *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.* 186, 1304–1310. doi: 10.1128/JB.186.5.1304-1310.2004

Troy, S. M., Duthie, C. A., Hyslop, J. J., Roehe, R., Ross, D. W., Wallace, R. J., et al. (2015). Effectiveness of nitrate addition and increased oil content as methane mitigation strategies for beef cattle fed two contrasting basal diets. *J. Anim. Sci.* 93, 1815–1823. doi: 10.2527/jas.2014-8688

Uehara, T., and Park, J. T. (2008). Growth of *Escherichia coli*: significance of peptidoglycan degradation during elongation and septation. *J. Bacteriol.* 190, 3914–3922. doi: 10.1128/JB.00207-08

Uehara, T., Parzych, K. R., Dinh, T., and Bernhardt, T. G. (2010). Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J.* 29, 1412–1422. doi: 10.1038/emboj.2010.36

United Nations–Department of Economic and Social Affairs/Population Division (2017). *World Population Prospects: the 2017 revision. Key findings and advance tables, ed*. United Nations New York: United Nations.

Van Soest, P. J. (1994). *Nutritional ecology of the ruminant*. 2nd. New York: Cornell University Press.

Vigors, S., O'Doherty, J. V., Kelly, A. K., O'Shea, C. J., and Sweeney, T. (2016). The Effect of Divergence in feed efficiency on the intestinal microbiota and the intestinal immune response in both unchallenged and lipopolysaccharide challenged ileal and colonic explants. *PLoS One* 11, e0148145. doi: 10.1371/journal.pone.0148145

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 839. doi: 10.1186/s12864-015-2032-0

Warren, M. J., Raux, E., Schubert, H. L., and Escalante-Semerena, J. C. (2002). The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* 19, 390–412. doi: 10.1039/b108967f

Wang, A., Gu, Z., Heid, B., Akers, R. M., and Jiang, H. (2009). Identification and characterization of the bovine G protein-coupled receptor GPR41 and GPR43 genes. *J. Dairy Sci*. 92, 2696–2705. doi: 10.3168/jds.2009-2037

Zapun, A., Contreras-Martel, C., and Vernet, T. (2008). Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiol. Rev.* 32, 361–385. doi: 10.1111/j.1574-6976.2007.00095.x

Zgurskaya, H. I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. *Front. Microbiol.* 2, 1–13. doi: 10.3389/fmicb.2011.00189

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lima, Auffret, Stewart, Dewhurst, Duthie, Snelling, Walker, Freeman, Watson and Roehe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 August 2019
doi: 10.3389/fgene.2019.00672

Comparative Transcriptomic and Proteomic Analyses Identify Key Genes Associated With Milk Fat Traits in Chinese Holstein Cows

*Chenghao Zhou†, Dan Shen†, Cong Li, Wentao Cai, Shuli Liu, Hongwei Yin, Shaolei Shi, Mingyue Cao and Shengli Zhang**

Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture & National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China

Milk fat is the most important energy substance in milk and contributes to its quality and health benefits. However, the genetic mechanisms underlying milk fat synthesis are not fully understood. The development of RNA sequencing and tandem mass tag technologies has facilitated the identification of eukaryotic genes associated with complex traits. In this study, we used these methods to obtain liver transcriptomic and proteomic profiles of Chinese Holstein cows (n = 6). Comparative analyses of cows with extremely high vs. low milk fat percentage phenotypes yielded 321 differentially expressed genes (DEGs) and 76 differentially expressed proteins (DEPs). Functional annotation of these DEGs and DEPs revealed 26 genes that were predicted to influence lipid metabolism through insulin, phosphatidylinositol 3-kinase/Akt, mitogen-activated protein kinase, 5′ AMP-activated protein kinase, mammalian target of rapamycin, and peroxisome proliferator-activated receptor signaling pathways; these genes are considered as the most promising candidate regulators of milk fat synthesis. The findings of this study enhance the understanding of the genetic basis and molecular mechanisms of milk fat synthesis, which could lead to the development of cow breeds that produce milk with higher nutritional value.

Keywords: milk fat, transcriptomic, proteomic, Chinese Holstein, liver

OPEN ACCESS

Edited by:
David E MacHugh, University College Dublin, Ireland

Reviewed by:
Kate Keogh, Teagasc Grange Animal and Bioscience Research Department (ABRD), Ireland
Tao Zhou, Auburn University, United States

***Correspondence:**
Shengli Zhang
zhangslcau@cau.edu.cn

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *30 January 2019*
Accepted: *27 June 2019*
Published: *13 August 2019*

Citation:
Zhou C, Shen D, Li C, Cai W, Liu S, Yin H, Shi S, Cao M and Zhang S (2019) Comparative Transcriptomic and Proteomic Analyses Identify Key Genes Associated With Milk Fat Traits in Chinese Holstein Cows. Front. Genet. 10:672. doi: 10.3389/fgene.2019.00672

INTRODUCTION

Milk products are an important part of our daily diet. There are a multitude of different milk products that vary in terms of composition, including fatty acid and protein content. Milk contains approximately 3–5% fat, which is the most important energy-rich substance it contains. The nutritional value of milk fat depends on the composition of fatty acids (FAs), which are classified according to hydrocarbon chain length as short-chain (C4–C10), medium-chain (C11–C17), and long-chain (LC, ≥C18) FAs, and according to the degree of saturation of the hydrocarbon chains as saturated (S)FAs, monounsaturated FAs, and polyunsaturated (PU)FAs. High concentrations of SFAs such as myristic acid (C14:0), lauric acid (C12:0), and palmitic acid (C16:0) increase low-density lipoprotein (LDL) concentration in the blood, which has been linked to cardiovascular and cerebrovascular diseases (Mensink et al., 2003). Meanwhile, PUFAs such as conjugated and unconjugated linoleic acid (C18:2) play a beneficial role in reducing blood lipids, suppressing the immune response, promoting bone formation, and stimulating lipid metabolism (Belury, 2002).

The ratio of PUFA to SFA is an important indicator of diet quality. The main proteins in milk are α_{s1}-casein (CN), α_{s2}-CN, β-CN, κ-CN, α-lactalbumin, and β-lactoglobulin, which are known to contribute to lipid synthesis and metabolism in humans (Mcgregor and Poppitt, 2013). The liver is a complex digestive gland in ruminant animals, including dairy cattle, and plays an important role in the metabolism of carbohydrates, fats, proteins, vitamins, hormones, and other substances. Nutrients absorbed from the digestive tract pass through the liver, enter the circulatory system, and finally arrive in the mammary glands of dairy cattle. The liver thus plays a critical role during lactation in cattle (Dorland et al., 2009; Graber et al., 2010; Schlegel et al., 2012).

There are few reports on the breeding of transgenic dairy cows, and cows that produce low-fat, high-protein milk have not been developed to date. The main constraint is the difficulty in obtaining animals that are true-breeding for this particular trait owing to the lack of information on related genes. The 29 autosomes of cows harbor most of the genes controlling milk traits and production (e.g., fat and protein content), including diacylglycerol O-acyltransferase (*DGAT*)*1* p.Lys232Ala and stearoyl-coenzyme (Co)A desaturase1 p.Ala293Val (Mele et al., 2007; Schennink et al., 2008; Conte et al., 2010), and many important or suggestive genomic regions have been identified (Schennink et al., 2009b; Stoop et al., 2009).

The development of RNA sequencing (RNA-seq) and tandem mass tag (TMT) technologies has enabled the identification of eukaryotic genes associated with complex traits through analysis of transcriptomic and proteomic profiles with low bias, broad dynamic range, low rate of false positive signals, and high reproducibility. The TMT method employs a set of amine-reactive isobaric tags to derivatize peptides at the N terminus and at lysine side chains, thereby facilitating simultaneous protein identification and quantification *via* mass spectrometry (MS) analysis of peptide fragments. Both RNA-seq and TMT have been widely used to screen for functional genes associated with milk composition in cattle and other domestic animals. In the present study, we compared the liver transcriptome and proteome profiles of Chinese Holstein cows with extremely high and low phenotypic values for milk fat and identified genes and proteins involved in milk production.

MATERIALS AND METHODS

Sample Collection

Based on milk production in their previous lactation, six Chinese Holstein cows—of which three were in their second and three in their third lactation—were selected from the Beijing Sanyuan Lvhe Dairy Farm and divided into high milk fat percentage (HP) and low milk fat percentage (LP) groups, each with three cows. The average milk fat percentage in this population was 3.7% (2.3–3.9%). Based on Dairy Herd Improvement system (DHI) data, we defined a high milk fat percentage group as those cows with 3.7% milk fat, and the low milk fat percentage group was composed of cows with 3.2% milk fat. The phenotype information of six Chinese Holstein cattle are showed in **Table S1**. The cows were kept in free stall housing, fed a total mixed ration (TMR, containing 16.1% crude protein, 22.9% acid detergent fiber), and had access to water ad libitum. Cows were milked three times daily in the milking parlor. The age differences among the cows in the second lactation were less than 45 days. Among the six cows, there were two pairs of half sibs consisting of one HP and one LP cow; the other two cows—one HP and one LP cow—were non-sibs. The cows were killed by electroshock, bled, skinned, and dismembered in the same slaughterhouse. Liver tissue samples (approximately 0.5–1.0 g) from each individual were removed within 30 min after slaughter. Five pieces of liver tissue samples per cow were carefully collected for RNA isolation, placed into a clean RNAse-free Eppendorf tube, and stored in liquid nitrogen. All sample collection procedures were carried out in strict accordance with the protocol approved by the Animal Welfare Committee of China Agricultural University (Permit Number: DK996).

RNA Isolation, Library Preparation, and Sequencing

Total RNA was extracted from the bovine liver tissue using the Trizol method (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. RNA degradation and contamination was monitored on 1% agarose gels. RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). The six purified RNA samples had an RIN ≥ 7.0, and a total of 1 μg RNA per sample was used as input material for RNA sample preparation. Sequencing libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×). First strand cDNA was synthesized using random hexamer primer and Moloney murine leukemia virus reverse transcriptase; second strand cDNA synthesis was then performed using DNA polymerase I and RNase H. Remaining overhangs were converted into blunt ends through exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with a hairpin loop structure was ligated to prepare the fragments for hybridization. In order to select cDNA fragments with a length of approximately 240 bp, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, MA, USA). A 3-μl volume of USER Enzyme (New England Biolabs) was incubated with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C. PCR was performed with Phusion High-Fidelity DNA polymerase, universal PCR primers, and index (X) primer. PCR products were purified (AMPure XP system), and library

quality was assessed with the Agilent Bioanalyzer 2100. Index-coded samples were clustered on a cBot Cluster Generation System using TruSeq PE Cluster Kit v.4-cBot-HS (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. After cluster generation, cDNA libraries were sequenced on an Illumina platform, and paired-end reads were generated.

Mapping and Annotation of Sequencing Reads

Raw data (raw reads) in fastq format were first processed with in-house Perl scripts. In this step, clean data (reads) were obtained by removing those containing adapter and poly-N sequences and low-quality reads. At the same time, Q20, Q30, GC content, and sequence duplication level of the clean data were calculated. All downstream analyses were based on high-quality clean data; these reads were mapped to the reference genome sequence (UMD3.1.80). Only reads with a perfect match or one mismatch were further analyzed and annotated based on the reference genome. HISAT2 (https://ccb.jhu.edu/software/hisat2/index.shtml) was used for mapping to the reference genome.

Quantification and Differential Gene Analysis by RNA-seq

Fragments per kilobase of exon per 106 mapped fragments (FPKM) values obtained using Cufflink v.2.1.1 software (http://cole-trapnell-lab.github.io/cufflinks/) were used as values for normalized gene expression. Differential expression analyses of HP vs. LP were performed using DESeq2 (Love et al., 2014), which provides statistical tools for identifying differential expression in digital gene expression data using a model based on the negative binomial distribution. The resultant p values were adjusted using Hochberg method for controlling the false discovery rate (FDR). q value < 0.01 and | log2 [fold change (FC)]| ≥ 1 were set as thresholds for significantly different expression.

Protein Isolation, Enzymolysis, and TMT Labeling

The 500 µl SDT buffer was added to the 50 mg samples, which were transferred to 2-ml tubes containing quartz sand (with 1/4-inch ceramic beads included for tissue samples). The lysate was homogenized twice for 60 s each (24 × 2, 6.0 m/s) with a homogenizer (MP Biomedicals, Solon, OH, USA). The homogenate was boiled for 3 min and then sonicated for 2 min. After centrifugation at 20,000 × g for 20 min at 4°C, the concentration of proteins in the filtrate was quantified with a BCA Protein Assay Kit (Bio-Rad, Hercules, CA, USA). DTT and UA buffer (8 M Urea, 150 mM Tris-HCl, pH 8.0) were added to 300 µg of the supernatant and the resulting mix was passed through a 10 KD filter. The protein samples were centrifuged with UA buffer, IAA (50mM IAA in UA), and NH_4HCO_3 buffer and then treated overnight with trypsin at a trypsin-to-protein ratio of 1:100. The peptide mixture (100 µg) of each sample was labeled using 10PLEX TMT reagent according to the manufacturer's instructions (Thermo Fisher Scientific).

The peptide mixture was loaded onto a reversed-phase trap column (Thermo Scientific Acclaim PepMap100, 100 µm × 2 cm, nanoViper C18) connected to a C18 reversed-phase analytical column (length = 10 cm, inner diameter = 75-µm, 3-µm resin; Thermo Fisher Scientific) in buffer A (0.1% formic acid) and separated for 1.5 h with a linear gradient of buffer B (98% acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min controlled by IntelliFlow technology (4%–7% buffer B for 2 min, 7%–20% buffer B for 65 min, 20%–35% buffer B for 12 min, 35%–90% buffer B for 2 min, and holding in 90% buffer B for 9 min).

Liquid Chromatography Tandem MS (LC-MS/MS) Analysis

LC-MS/MS analysis was performed on a Q-exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to an Easy nLC chromatograph (Proxeon Biosystems, now Thermo Fisher Scientific) for 90 min. The instrument was operated in positive ion mode. MS data were acquired using a data-dependent top 10 method to dynamically select the most abundant precursor ions from the survey scan (300–1,800 m/z) for higher-energy collisional dissociation (HCD) fragmentation. The automatic gain control target was set to 1e6, with a maximum injection time of 50 ms. The duration of dynamic exclusion was 40.0 s. Survey scans were acquired at a resolution of 70,000 at m/z 200, and resolution for HCD spectra was set to 35,000 at m/z 200 (TMT 10PLEX), with an isolation window of 1.6 Th. Normalized collision energy was 35 eV, and the underfill ratio—which specifies the minimum percentage of the target value likely to be reached at maximum fill time—was defined as 0.1%. The instrument was run with the peptide recognition mode enabled.

Database Search and Protein Identification and Quantification

For peptide identification and quantification, MS/MS data were searched against the "Uniprot-Bos taurus_32310_20180905.fasta" file using Maxquant version 1.6.0.16. The following parameters were used: trypsin as enzyme specificity; maximum two missed cleavages permitted; fixed modification: carbamidomethylation of cysteine residues; variable modifications: oxidation of methionine residues and N-terminal acetylation; first search peptide tolerance of 20 ppm; main search peptide tolerance of 4.5 ppm. Protein quantification was based on the razor and unique peptides. Fold decrease/increase >1.2 and $p < 0.05$ were set as the threshold for identifying differentially expressed proteins (DEPs).

Gene Ontology (GO) Enrichment Analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment Analysis

GO enrichment analysis of differentially expressed genes (DEGs) was performed with the GOseq R packages based on a Wallenius non-central hyper-geometric distribution (Young

et al., 2010), which can adjust for gene length bias. The KEGG database (Kanehisa et al., 2007) is used to analyze high-level functions of a biological system based on molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput approaches (http://www.genome.jp/kegg/). We used KOBAS software (Mao et al., 2005) to assess the enrichment of DEGs in KEGG pathways.

Protein–Protein Interactions (PPIs) and Mapping of Quantitative Trait Loci (QTL)

The sequences of DEGs were searched against the genome of a related species using blastx; we searched the STRING database (http://string-db.org/) to determine the predicted PPIs of these DEGs. The PPIs were visualized using Cytoscape (Shannon et al., 2003). We also integrated the DEGs and QTL for milk fat traits from the QTLdb database (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) into our analysis.

Verification of RNA-seq and TMT Data

RT-qPCR primers were designed to span the exon-exon boundaries of eight genes selected by RNA-seq; RT-PCR analysis was performed using a SYBR® Premix Ex Taq™ II (Tli RNaseH Plus), ROX plus (RR82LR, TaKaRa) on a ABI7500 Real-Time PCR Detection System (Applied Biosystems), according to the manufacturer's instructions.

The protein expression levels obtained using TMT analysis were confirmed by quantifying the expression levels of five selected proteins by a parallel reaction monitoring (PRM) analysis carried out at the Beijing Bangfei Bioscience Co., Ltd. (Beijing, China). PRM is a targeted method of quantification performed using high-resolution hybrid mass spectrometers such as quadrupole-Orbitrap (q-OT). Signature peptides for the target proteins were defined according to the TMT data, and only unique peptide sequences were selected for the PRM analysis. Each protein sample (50 μg) was separated using a nanoliter flow HPLC liquid phase system Easy nLC 1200 (Thermo Fisher). Samples were loaded by an autosampler into a mass spectrometer pre-column C18 trap column (C18, 3 μm, 100 μm × 20 mm) and separated by an analytical column C18 column (C18, 3 μm, 75 μm × 150 mm). After peptide separation, targeted PRM mass spectrometry was performed using a Q-Exactive Plus mass spectrometer (Thermo Scientific). The result of mass spectrum was analyzed using the software Skyline 4.1.

RESULTS

Overview of RNA Transcriptomic Profiles of Cow Liver Tissue

A total of 197,358,565 paired-end reads were obtained by RNA-seq. The quality value of Q30 for sequencing was no less than 96.40% for each sample. An average of 91.55% (range: 93.87%–95.31%) of reads were mapped to the bovine genome (Ensembl UMD3.1) using HISAT2. Of these, approximately 91.42% (range: 90.75%–92.10%) were uniquely mapped and 3.29% (range: 3.07%–3.84%) were multi-mapped reads (**Table S2**). Additionally, of the total mapped reads, roughly 70% in each group corresponded to exons (**Figure S1**).

Analysis of DEGs

The expression levels of known and novel genes were calculated as FPKM using DESeq2 (Love et al., 2014), which provides statistical approaches for identifying differentially expressed known and novel genes based on a negative binomial distribution model. A total of 23,098 genes were expressed in liver tissue. Pairwise comparisons according to stringent criteria—i.e., $|\log2\ (FC)| > 2$ and $q < 0.01$—were carried out to identify DEGs (**Figure 1**). A total of 321 genes were differentially expressed between HP and LP groups, including 117 that were up-regulated and 204 that were down-regulated (**Table S3** and **Figure 1**, q value < 0.01). The results of cluster analysis of DEGs are depicted in a heatmap (**Figure S2**).

Functional Analysis of DEGs

We used GOseqR packages and the KEGG database to determine the function of the identified DEGs. The top three functions related to metabolism were "cell adhesive protein binding involved in bundle of His cell-Purkinje myocyte communication," "polyamine oxidase activity," and "serine and oxidoreductase activity, acting on the CH-NH group of donors, oxygen as acceptor" (KS ≤ 1.0E−30) (**Table S4**). We identified a metabolic network comprising 22 DEGs involved in insulin production (mitogen-activated protein kinase [*MAPK*]*9*, cyclic AMP response element-binding protein [*CREB*]*1*, protein phosphatase 1 regulatory subunit [*PPP1R*]*3C*, nuclear factor κB inhibitor α [*NFKBIA*], peroxisome proliferator-activated receptor γ, coactivator 1α [*PPARGC1A*], and forkhead box [*FOX*]*O1*); insulin resistance (*MAPK9*, *CREB1*, *PPP1R3C*, *NFKBIA*, *PPARGC1A*, and *FOXO1*); phosphatidylinositol 3-kinase (PI3K)/Akt signaling (DNA damage-inducible transcript [*DDIT*]*4*, *CREB1*, G protein subunit γ [*GNG*]*7*, platelet-derived growth factor subunit [*PDGF*]*A*, ephrin A1, protein kinase [*PKN*]*2*, breast cancer type 1 susceptibility gene, and tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein η); MAPK signaling (FBJ murine osteosarcoma viral oncogene homolog[*FOS*], *MAPK9*, growth arrest and DNA damage-inducible α [*GADD*]*45A*, dual specificity phosphatase [*DUSP*]*1*, platelet-derived growth factor subunit A [*PDGFA*], *MAP4K3*, *GADD45B*, and *DUSP8*); prolactin production (*FOS*, suppressor of cytokine signaling [*SOCS*]*1*, *MAPK9*, *CREB1*, and *SOCS2*); 5′ AMP-activated protein kinase (AMPK) signaling (hepatocyte nuclear factor 4α [*HNF*]*4A*, *PPARGC1A*, and *FOXO1*); mammalian target of rapamycin (mTOR) signaling (DNA damage inducible transcript [*DDIT*]*4* and disheveled segment polarity protein 2); and PPAR signaling (*PPARδ* [*PPARD*]) (**Table S5** and **Figure 2**). The FOXO and CREB families have key features for the integration of insulin production and insulin resistance signaling with glucose and lipid metabolism (Lee and Dong, 2017). The PI3K/AKT signaling pathway played a key role in regulating lipid metabolism in lactating goats (Li et al., 2018). The results of a previous GWAS study showed that the MAPK signaling pathway was overrepresented for milk protein and fat content (Cecchinato et al., 2019). Prolactin production, AMPK, and PPAR signaling pathways are well known for regulating milk

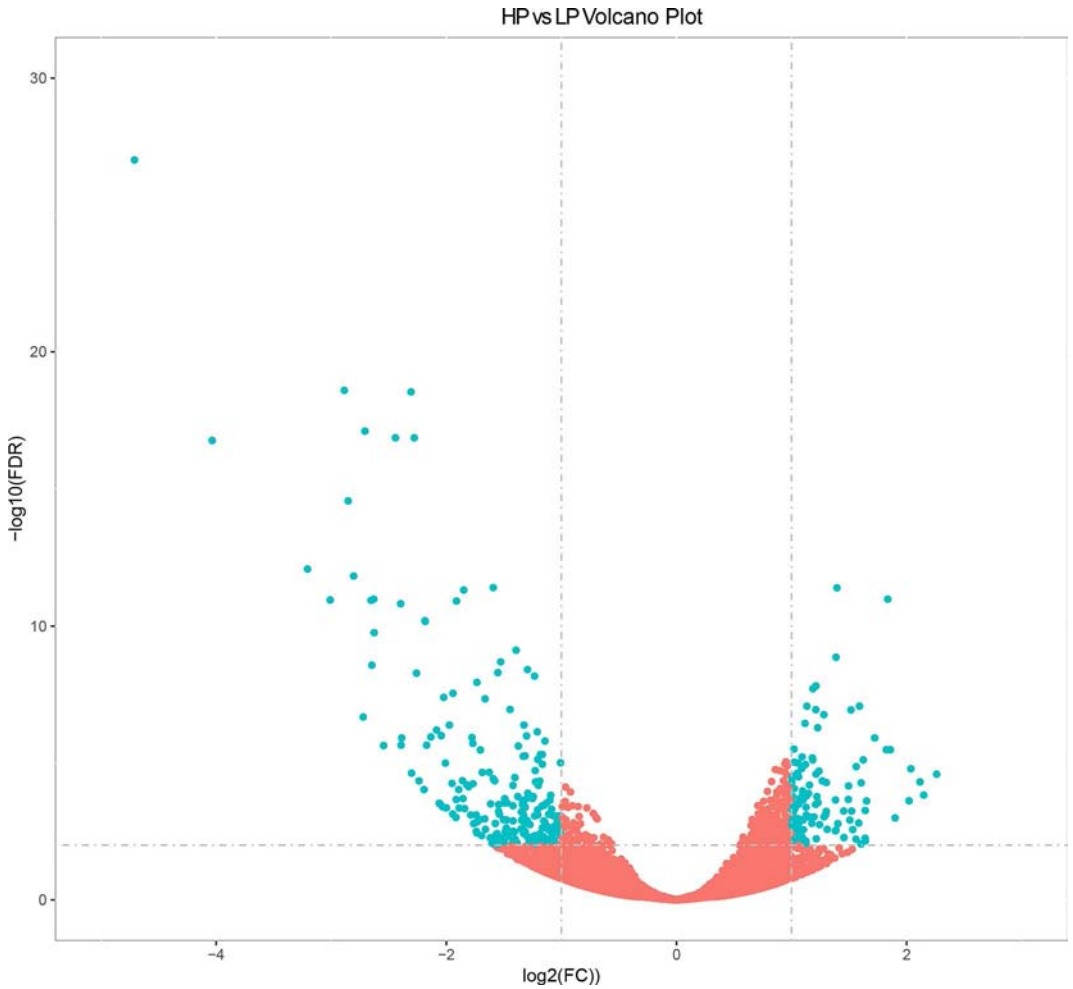


FIGURE 1 | Volcano plot displaying differential expressed genes in bovine liver tissues with transcriptomic analyses within two different comparison groups. The *y* axis corresponds to the mean expression value of log10 (*q* value), and the *x* axis displays the log2 fold change value. The blue dots represent the significantly differential expressed transcripts; the red dots represent the transcripts whose expression levels did not reach statistical significance.

fat synthesis (Schennink et al., 2009a; Liu et al., 2016; Gao et al., 2017). These networks play critical roles in the regulation of milk fat synthesis (Anderson et al., 2007; Bionaz and Loor, 2011). On the basis of their biological function and PPI analysis, 16 of the 22 genes were considered as important for lipid metabolism in the liver (**Table 1**).

Protein Identification and Quantification by TMT

A total of 112,916 spectra were obtained in the 10PLEX LC-MS/MS analysis. After pooling samples from the two groups, 31,327 unique peptides were identified, including 4,356 proteins that were originally identified with the Q-exactive Plus Orbitrap mass spectrometer (**Figure S3a**). To eliminate false positives, we controlled FDR to 1% at both the peptide and protein levels using the MaxQuant reversed sequence database. The number of proteins identified at various molecular weight ranges were as follows: 0–50 kDa, 2,349; 50–100 kDa, 1,215; 100–150 kDa, 350; 150–200 kDa, 100; 200–300 kDa, 86; and 300–3850 kDa, 36. Collectively, these 4,136 proteins accounted for 94.95% of those identified (**Figure S3b**). In addition, most proteins had high peptide coverage; 85.61% and 14.39% had <50% and >50% sequence coverage, respectively (**Figure S3c**). Among the identified proteins, 48.39% were represented by fewer than five peptides (**Figure S3d**), indicating good sequence coverage. Information on the identification of proteins is shown in **Supplementary Tables S6, S7**.

Analysis of DEPs

Based on the selection criteria (fold decrease/increase >1.2 and $p < 0.05$), we identified 76 DEPs in the HP vs. LP comparisons, including 25 up-regulated and 51 down-regulated DEPs (**Table S7** and **Figure 3**). Clusters of all DEPs were visualized by a heatmap (**Figure S4**).

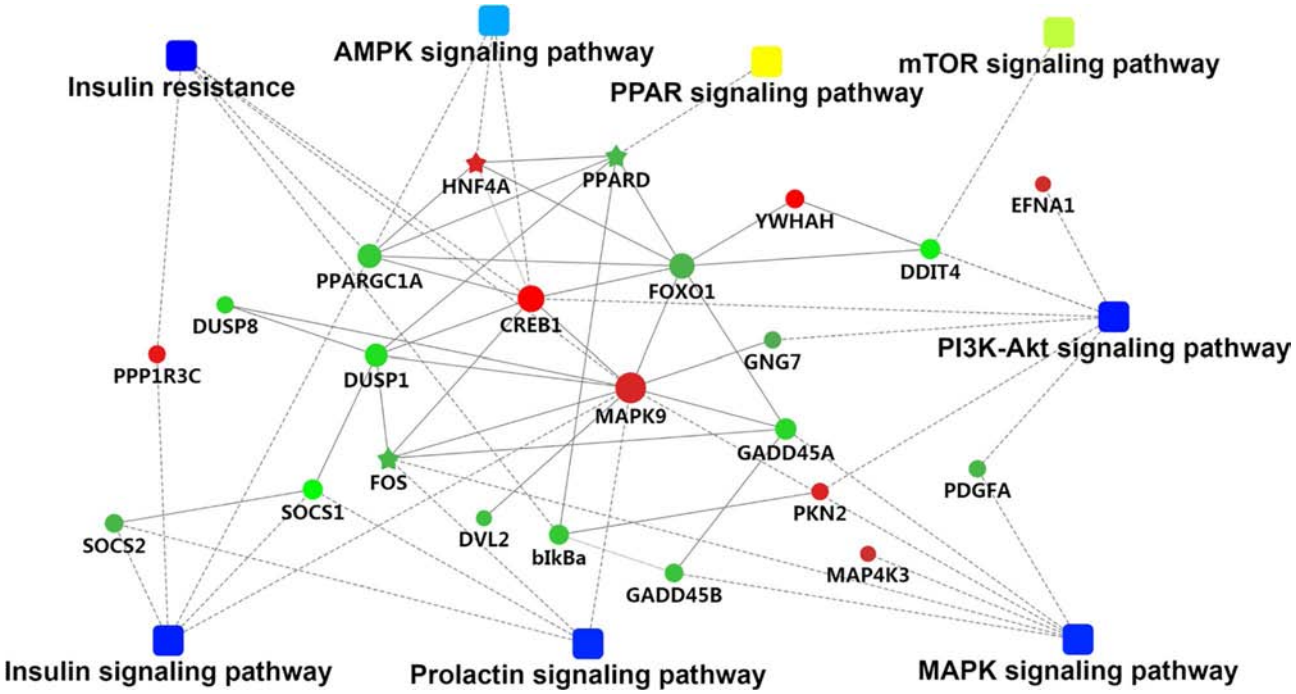


FIGURE 2 | The metabolic network comprising candidate genes in protein–protein interactions (PPI) network and pathways with transcriptomic analyses. The round nodes indicate genes, red indicates up-regulation, and green indicates down-regulation. The rectangular node represents the KEGG pathway/biological process, and the significant *p* value is represented by yellow-blue gradient; yellow indicates a small *p* value, while blue indicates a large *p* value.

TABLE 1 | Expression changes of the candidate genes in bovine liver tissue with transcriptomic analyses.

Symbol	Chromosome	Gene name	HP counts	LP counts	Log2 fold change	*Q*-value
SOCS1	25	Suppressor of cytokine signaling 1	163	2775	–3.0090	1.11E-11
SLCO4A1	13	Solute carrier organic anion transporter family member 4A1	2616	21071	–2.7070	7.91E-18
DDIT4	28	DNA damage inducible transcript 4	1204	11321	–2.6543	1.14E-11
MTHFR	16	Methylenetetrahydrofolate reductase	6728	36841	–2.1832	6.73E-11
GADD45A	3	Growth arrest and DNA damage inducible alpha	2506	14092	–2.0076	1.01E-05
PPARGC1A	6	PPARG coactivator 1 alpha	1103	4745	–1.7676	5.69E-05
PDGFA	25	Platelet derived growth factor subunit A	55	158	–1.3385	6.25E-04
SOCS2	5	Suppressor of cytokine signaling 2	596	1650	–1.3137	4.15E-04
FOXO1	12	Forkhead box O1	1708	4036	–1.2051	7.46E-06
SLC22A1	9	Solute carrier family 22 member 1	18901	47883	–1.1927	7.20E-03
MAPK9	7	Mitogen-activated protein kinase 9	1898	788	1.1277	1.63E-04
HNF4A	13	Hepatocyte nuclear factor 4 alpha	11382	4804	1.1560	6.41E-05
CREB1	2	cAMP responsive element binding protein 1	667	175	1.6049	5.35E-05
SYBU	14	Syntabulin	2034	560	1.6236	7.72E-06
HNF4G	14	Hepatocyte nuclear factor 4 gamma	2881	642	1.8243	3.28E-06

Functional Analysis of DEPs

To assess the biological significance of these DEPs in hepatic tissue of Holstein cows with different milk fat compositions, the DEPs were further classified based on GO and KEGG functional annotations. For the "cellular component" aspect, the classification analysis revealed that most of the DEPs were related to mitochondria (33.90%), with four related to the endoplasmic reticulum (chloride channel CLIC-like protein [*CLCC*]*1*, heat shock 70 kDa protein [*HSPA*]*13*, transmembrane protein [*TMEM*]*33*, and solute carrier family 27 member 2 [*SLC27A2*]). For "biological process," the GO terms were mainly associated with "oxidation-reduction process" (methionine-R-sulfoxide reductase [*MSR*]*B1*, ATP binding cassette subfamily D member 3 [*ABCD3*], aldehyde dehydrogenase 7 family member A1 [*ALDH7A1*], AUH protein [*AUH*], SLC27A2, electron transfer flavoprotein subunit alpha [ETFA], ENSBTAG00000000229, proline dehydrogenase 1 [*PRODH*], NAD-dependent protein deacetylase [*SIRT3*], succinate dehydrogenase cytochrome b560 subunit [*SDHC*], retinol dehydrogenase [*RDH*]*13*, hydroxysteroid (17-beta) dehydrogenase [*HSD17B*]*13*, succinate–CoA ligase [ADP/GDP-forming] subunit alpha [*SUCLG1*], NADH dehydrogenase

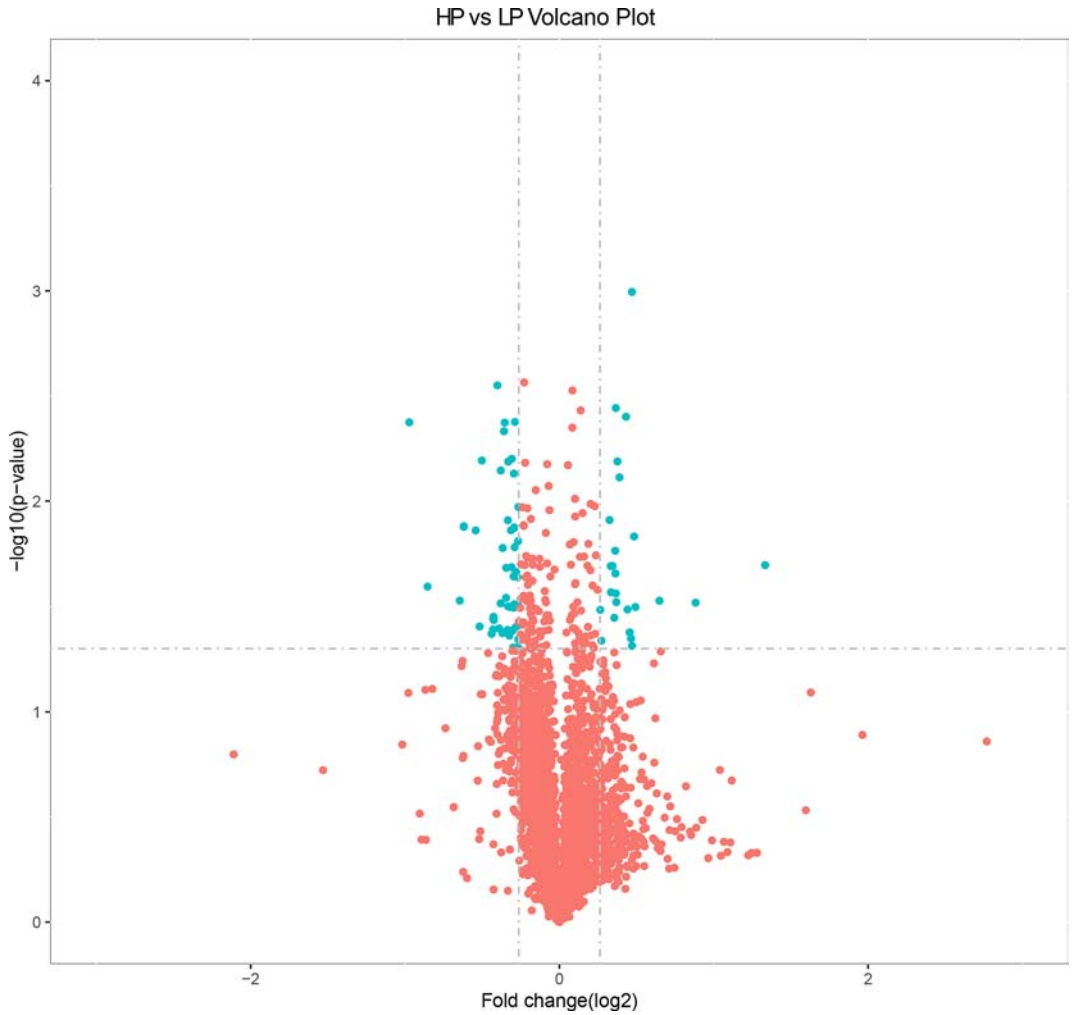


FIGURE 3 | Volcano plot displaying differential expressed proteins in bovine liver tissues with proteomic analyses within two different comparison groups. The *y* axis corresponds to the mean expression value of log10 (*p* value), and the *x* axis displays the log2 fold change value. The blue dots represent the significantly differential expressed transcripts; the red dots represent the transcripts whose expression levels did not reach statistical significance.

[ubiquinone] 1 [*NDUF*]*B3*, and *NDUFA2*), "monocarboxylic acid catabolic process" (alanine–glyoxylate aminotransferase [*AGXT*]*2*, *ABCD3*, *AUH*, *SLC27A2*, and *ETFA*), "fatty acid beta-oxidation" (*ABCD3*, *AUH*, *SLC27A2*, and *ETFA*), "carboxylic acid catabolic process" (*ABCD3*, *AUH*, *SLC27A2*, *ETFA*, *PRODH*, and *AGXT2*), "fatty acid catabolic process" (*ABCD3*, *AUH*, *SLC27A2*, and *ETFA*), and "very long-chain fatty acid catabolic process" (*ABCD3* and *SLC27A2*). Functional associations among these GO terms were visualized using the STRING database (**Table S8** and **Figure 4**). We also found potentially relevant GO terms in "molecular function," including "fatty acid transporter activity" (*ABCD3* and *SLC27A2*); "long-chain fatty acid binding" (S100 calcium-binding protein [*S100*] *A9* and *S100A8*); and "oxidoreductase activity" (*MSRB1*, *ALDH7A1*, *ETFA*, *ENSBTAG00000000229*, *PRODH*, *SDHC*, *NDUFA2*, *RDH13*, and *HSD17B13*).

KEGG pathway analysis of the significantly altered proteins revealed 13 enriched canonical pathways ($p < 0.05$) (**Table S9**); the top three related to metabolism were "oxidative phosphorylation," "citrate cycle and glycine," and "serine and threonine metabolism" (p = 3.11E−04, 2.45E−03, and 4.76E−03, respectively). The major functional associations within these pathways were visualized using the STRING database. Notably, five genes encoding enriched DEPs were related to insulin resistance (*SLC27A2* and phosphoenolpyruvate carboxykinase [*PCK*]*1*), insulin secretion (phosphoinositide phospholipase C-β2 [*PLCB2*]), insulin signaling (*PCK1*), PI3K/Akt signaling (*PCK1*), AMPK signaling (*PCK1*), and PPAR signaling (cytochrome P450 family 4 subfamily A member [*CYP4A*]*11*, *SLC27A2*, carnitine palmitoyltransferase [*CPT*]*2*, and *PCK1*). The details of the five candidate genes with proteomic profiles are shown in **Table 2**.

Validation of DEGs and DEPs

To validate the accuracy of the DEGs detected by RNA-seq analysis, we used real-time reverse transcription-quantitative polymerase chain reaction (RT-qPCR) to evaluate the expression

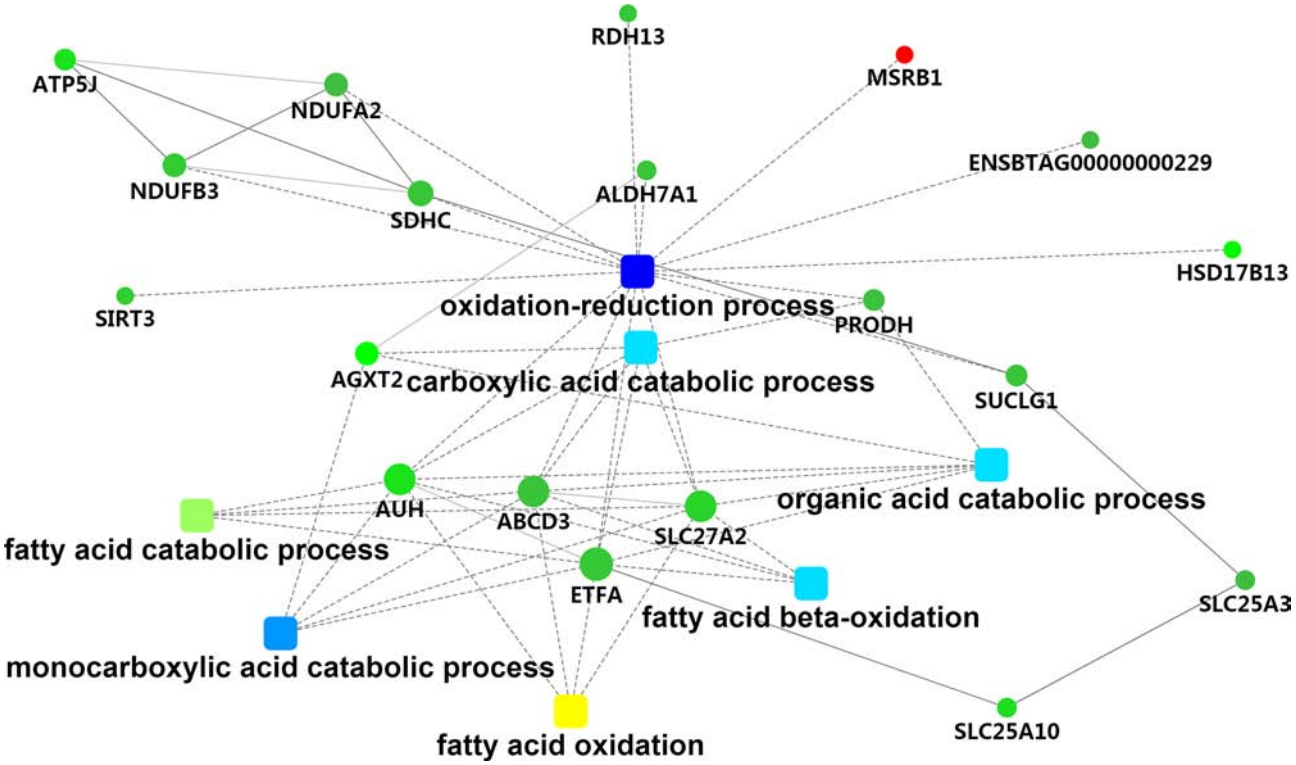


FIGURE 4 | The metabolic network comprising candidate genes in PPI network and GO terms with transcriptomic analyses. Round nodes indicate genes, red indicates up-regulation, and green indicates down-regulation. The rectangular node represents the KEGG pathway/biological process, and the significant *p* value is represented by yellow-blue gradient; yellow indicates a small *p* value, while blue indicates a large *p* value.

TABLE 2 | Expression changes of the candidate genes in bovine liver tissue with proteomic analyses.

Symbol	Protein ID	Gene name	Chromosome	Sequence length	Log2 fold change	*P*-value
SLC27A2	F1MQP2	Solute carrier family 27 (fatty acid transporter), member 2	10	620	–0.3590	0.0047
PCK1	F1N1Z7	phosphoenolpyruvate carboxykinase 1	13	622	–0.4296	0.0406
CPT2	F1N1M7	carnitine palmitoyltransferase 2	3	658	–0.8542	0.0253
SIRT3	G5E521	sirtuin 3	11	333	–0.3330	0.0315
CYP4A11	F1ME58	cytochrome P450, family 4, subfamily A, polypeptide 11	3	514	–0.2852	0.0396
PLCB2	E1B7M6	Phospholipase C beta 2	10	1171	0.3344	0.0270

levels of eight DEGs: *PPARGC1A*, *DDIT4*, suppressor of *SOCS1*, solute carrier family 22 member 1 [*SLC22A1*], *HNF4A*, *PDGFA*, syntabulin [*SYBU*], and *MAPK9*. The expression levels of these genes in each group are shown in **Figure S6**. The eight genes selected were differentially expressed among the HP vs LP comparison group, and the RNA-seq data were concordant with those obtained by RT-qPCR (**Figure S6**).

The PRM assay was used to confirm the identity of several DEPs identified in the TMT analysis. Parallel reaction monitoring (PRM) technology uses a quadruple mass analyzer to selectively detect target proteins and target peptides. This technology has higher specificity and sensitivity than selected reaction monitoring (SRM) technology. As this assay requires the signature peptide of the target protein to be unique, we only selected proteins with a unique signature peptide sequence for the PRM analysis. Five DEPs (up-regulated:carnitine *CPT2*; down-regulated: Cytochrome [cyt b, *RDH13*, *CYP4A11*, and *SLC27A2*) were selected for the PRM analysis (**Figure S7**).

Integrated Analysis of DEGs and DEPs From TMT and RNA-Seq Data

The Pearson correlation coefficient for the log2 function of HP vs. LP was 0.31, indicating that mRNA and protein levels were only partially correlated overall. Only *SLC22A1* and Heat shock protein family A member 13 [*HSPA13*] were identified as both DEGs and DEPs. On the basis of these results, we propose that post-transcriptional regulatory activity contributes to milk fat lipid anabolism.

Integrated Analysis of DEGs, DEPs, and Animal QTLdb

We integrated DEGs and QTL for milk production traits from the QTLdb database that were detected either by QTL mapping studies or genome-wide association studies (GWAS) by comparing their chromosome positions in order to gain further insight into the association between DEGs and milk fat traits. For QTL detected by QTL mapping studies, only those with a confidence interval less than 1 Mb were considered as a QTL region; for those identified by GWAS, the 200 kb up-/downstream of significant single nucleotide polymorphisms (SNPs) were defined as a QTL region. Among the DEGs and DEPs, 199 genes regions were located within or overlapped with QTL regions (**Figure S5**).

DISCUSSION

FAs in milk originate from two sources: some are synthesized *de novo* by mammary epithelial cells (MECs), including nearly all short (C4–C8) and middle chain (C10–C14) FAs and half of C16 FAs; the remaining C16 FAs and LCFAs (>C16) are obtained by MECs directly from the blood.

After rumen fermentation, digestive tract absorption, liver metabolism, and so on, compounds such as acetic acid, β-hydroxybutyric acid (BHBA), free (F) FAs, etc. from the absorption and conversion of dietary nutrients are used in the mammary gland to synthesize milk fat. The metabolism, transformation, and utilization of these precursors in the body directly affect milk fat content. In addition, acetic acid, BHBA, and FFAs act as signaling molecules to modulate lipid synthesis through a feedback mechanism in the liver and adipose tissue.

The lactation process of dairy cows has periodicity and can be divided into early non-lactating period, late non-lactating period, early lactation period, peak lactation period, middle lactation period, and late lactation period. The peak lactation period occurs 6–8 weeks after delivery. After the peak period to 30–35 weeks after delivery, it is the middle of lactation, and the milk yield in the medium term is slightly lower than that in the early stage; however, the milk components are relatively stable.

In this study, we identified genes associated with milk fat and milk FA production by examining the transcriptome and proteome profiles of liver tissue samples from Chinese Holstein cows with extremely high or low milk fat percentage. A comparative analysis revealed 321 DEGs and 76 DEPs; 8 DEGs of *PPARGC1A*, *DDIT4*, *SOCS1*, *SLC22A1*, *HNF4A*, *PDGFA*, *SYBU*, and *MAPK9* and 5 DEPs of *CPT2*, *cytb*, *RDH13*, *CYP4A11*, and *SLC27A2* were verified by RT-qPCR and PRM, respectively, and the results were consistent with the previous experiments, confirming the reliability of this multi-omics study. Some of the genes with known roles in milk production such as *DGAT1* (Grisart et al., 2004), growth hormone receptor[*GHR*] (Blott et al., 2003), and stearoyl-CoA desaturase[*SCD*] (Kinsella, 1972) did not differ between the two groups. It is likely that factors whose expression differed significantly between HP and LP cows have been fixed through long-term genetic selection. In particular, *SLC22A1* and *HSPA13* were identified as both DEGs and DEPs. A functional enrichment analysis identified for the first time 22 DEGs (*SLC22A1*, *MAPK9*, *PPARGC1A*, *FOXO1*, *SOCS1*, *SOCS2*, *CREB1*, *HNF4A*, *HNF4G*, *GADD45A*, *DUSP1*, *PDGF*, *SYBU*, *DDIT4*, BMP and activin membrane bound inhibitor [*BAMBI*], methylenetetrahydrofolate reductase [*MTHFR*], *SLC27A2*, *PCK1*, *CPT2*, *SIRT3*, *CYP4A11*, and *PLCB2*) as candidate genes that regulate milk fat synthesis, transport, and metabolism.

DEGs for Milk Fat Traits

MAPK9 encodes a member of the MAPK family. These proteins act as an integration point for multiple biochemical signals and are involved in a variety of cellular processes, including cell proliferation and differentiation, transcriptional regulation, and development. A previous study indicated that *MAPK9* is implicated in the response to intramammary challenge and negative energy balance in cows (Moyes et al., 2010). *MAPK9* is important in the proposed network of milk fat synthesis that includes encompassing MAPK [c-Jun N-terminal kinase (JNK)] and insulin signaling and insulin resistance (Liang et al., 2017). It is likely that the high expression of *MAPK9* is involved in integrating insulin production, insulin resistance, and MAPK and prolactin production signaling to increase lipid synthesis in the liver. *FOXO1* belongs to the forkhead family of transcription factors that are characterized by a distinct forkhead domain. The FOXO1/Akt pathway plays a critical role in gluconeogenesis in the liver (Yang et al., 2018). A previous transcriptome analysis of the liver suggested that *FOXO1* influences milk fat synthesis (Jacometo et al., 2016). In addition, the low expression levels of *FOXO1* suggested it may activate insulin production and insulin resistance signaling to enhance glucose and lipid metabolism in the liver. Suppressor of cytokine signaling (SOCS) family genes such as *SOCS1* and *SOCS2* encode signal transducer and activator of transcription (STAT)-induced STAT inhibitor proteins, which are cytokine-inducible negative regulators of cytokine signaling. A previous study reported several SNPs near the *SOCS1*, *SOCS3*, *SOCS5*, and *SOCS7* genes that were significantly associated with protein yield (Arun et al., 2015), suggesting that *SOCS1* and *SOCS2* interact with other genes to influence milk production and composition. CREB1 protein is phosphorylated by several protein kinases and induces gene transcription in response to hormonal stimulation *via* the cAMP pathway, leading to the regulation of lipid metabolism (Ikoma-Seki et al., 2015; Mucunguzi et al., 2017). *HNF4G* encodes HNF4γ, a nuclear transcription factor that binds DNA as a homodimer to control the expression of HNF1α, a transcription factor that regulates hepatic gene expression. *HNF4G* may also play a role in intramuscular fat deposition in beef cattle (Ramayo-Caldas et al., 2014), while the paralog *HNF4A* negatively regulates cholesterol metabolism and bile acid synthesis in the liver (Shirpoor et al., 2018). In the present study, the expression abundance of *HNF4G* and *HNF4A* in HP (2,881 and 11,382 reads) was threefold and twofold higher than LP groups (642 and 4,804 reads), respectively, revealing its high expression and importance in high milk fat percentage groups. *GADD45A* is up-regulated under stressful growth arrest conditions and by treatment with DNA-damaging agents. The protein encoded by this gene activates p38/JNK signaling *via*

MEKK4/MTK1 kinase, which is known to regulate fat deposition in pork (Cho et al., 2015). In the present study, the expression of *DUSP1*was fivefold higher in the LP group (13,073 reads) than in the HP group (2,380 reads), suggesting that high expression levels of *DUSP1* modulate lipid metabolism and synthesis (Nukitrangsan et al., 2011). *PDGF* belongs to the same protein family as vascular endothelial growth factors, which play an essential role in the regulation of embryonic development; cell proliferation, migration, and survival; and chemotaxis. PDGF is also involved in the synthesis of monounsaturated FAs in cells (De Brachene et al., 2017). *SYBU* encodes a microtubule-associated protein that mediates anterograde transport of vesicles to neuronal processes. *SYBU* is phosphorylated by exchange protein directly activated by cAMP (Epac)2 agonist 8-pCPT-2′-O-Me-cAMP (Ying et al., 2012). The findings of the present study suggest that a high expression level of *SYBU* is an effector of Epac2 that contributes to cAMP-induced insulin secretion as well as milk production and composition. *DDIT4* regulates cell growth, proliferation, and survival by suppressing the activity of mTOR complex 1, which is involved in the response to changes in cellular energy level and stress. *DDIT4* has been proposed to be a negative regulator of cell proliferation and cell growth in goat, thereby affecting the synthesis of milk fat (Crisà et al., 2016). *BAMBI* stimulates adipogenesis by suppressing carboxypeptidase A4—a negative regulator of adipogenesis that modulates local and systemic insulin sensitivity—through interactions with genes known to regulate milk production and composition (He et al., 2016). *MTHFR* catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate (nicotinamide adenine dinucleotide phosphate), a co-substrate for homocysteine remethylation to methionine. *MTHFR* gene deficiency may enhance liver injury by altering methylation capacity, inflammation, and lipid metabolism (Leclerc et al., 2018).

DEPs for Milk Fat Traits

FAs entering the liver are mainly derived from FFAs produced by body fat mobilization. Triglyceride (TG) in adipose tissue is hydrolyzed into FFAs and glycerol by hormone-sensitive lipase and released into the blood. FFAs form a complex with albumin that is absorbed and utilized by the liver; FA transport protein (FATP)2 encoded by *SLC27A2* is a transmembrane protein transporter involved in this uptake. FATP2 has very long-chain acyl-CoA synthetase activity and converts free (F) LCFAs into fatty acyl-CoA esters. Interestingly, FATP2 expression was 1.28-fold higher in LP compared to HP cows. This may be related to lipid accumulation in the liver, which reduces milk fat precursor production in the liver. FATP2 overexpression in the liver is related to hepatic steatosis (Krammer et al., 2011), which reflects increased accumulation of lipids (mainly TG) in hepatocytes. Although FATP2 promotes the uptake of LCFAs, the esterification rate of LCFAs in the liver is higher than its decomposition rate and the rate of very (V)LDL transport to remove TGs, which thus accumulates as the lipid concentration in the cytoplasm decreases. Indeed, high *SLC27A2* transcript levels in blood cells are associated with lower TG levels (Sanchez et al., 2012). However, platelet glycoprotein 4 [CD36]—another LCFA transporter—showed a tendency (albeit non-significant) towards overexpression in HP, suggesting that LCFA uptake occurs *via* distinct mechanisms. PCK1 catalyzes the irreversible formation of phosphoenolpyruvate from oxaloacetate in gluconeogenesis. In non-ruminant animals, *PCK1* expression is induced by starvation and decreases during feeding. It is repressed by insulin and induced by glucagon and glucocorticoids (Hanson and Reshef, 1997). However, in ruminants its expression is not related to feed restriction but is induced by increased feed intake and monensin feeding, leading to increased ruminal propionate production (Greenfield et al., 2000; Velez and Donkin, 2005; Karcher et al., 2007). *PCK1* promoter activity is linearly induced by propionate in bovine (Zhang et al., 2016). In the present study, the *PCK1* protein level was higher in the LP than in the HP group, indicating a higher rate of gluconeogenesis in the liver. Increased utilization of hepatic lactolipid synthesis precursors such as propionic acid to generate glucose enhances hepatic FA oxidation to CO_2, reflecting a redistribution of lactolipid precursors in the liver. *PCK1* is also involved in glyceroneogenesis, which catalyzes the production of glycerol-3-phosphate for FA esterification (Beale et al., 2007; Hosseini et al., 2015). Thus, the up-regulation of PCK1 in LP may be associated with TG accumulation in the liver. The increased level of acyl-CoA oxidase1—an enzyme involved in FA β-oxidation that stimulates ATP production to support gluconeogenesis and prevent lipid esterification and accumulation in the liver (Aoyama et al., 1994)—in LP suggests that lipid balance is regulated *via* modulation of glucose metabolism. Feeding strategies that increase rumen propionate production and thus induce *PCK1* expression are often used to meet increased glucose requirements and reduce the effects of fatty liver during early lactation in cows (Zhang et al., 2016). Consistent with our findings, cows with high liver fat content showed elevated expression of hepatic gluconeogenesis genes (Hammon et al., 2009), suggesting a higher gluconeogenic capacity in the liver. PPARγ is a known pro-adipogenic factor. We found here that the family with sequence similarity (FAM)120A—also known as bovine constitutive coactivator of PPARγ-like protein 1—was highly expressed in the LP group. PPAR signaling promotes the expression of the FA oxidation-related genes *CYP4A11* and *CPT2*, which is consistent with our results. We speculate that liver FA oxidation capacity is higher in the LP than in the HP group, leading to a reduction in milk fat synthesis precursors. β-Oxidation is the main pathway of FA catabolism. *CPT2* converts acylcarnitine translocated to the mitochondrial matrix into acyl-CoA and free carnitine and is a rate-limiting enzyme for the transport of LCFAs into mitochondria for β-oxidation (Isackson et al., 2013). Silent mating type information regulation 2 homolog (Sirt3) regulates FA β-oxidation in the liver *via* AMPK and sterol regulatory element-binding protein1, promoting FA utilization and thereby preventing fat heterotopia (Kong et al., 2016). *CPT2* and *Sirt3* proteins were highly expressed in the LP group, indicating enhanced FA β-oxidation. In addition to β-oxidation, microsomal ω-oxidation mediated by cytochrome P450 enzymes played a

key role in lipid synthesis and lipid accumulation (Hardwick, 2008). Through preferential hydroxylation of FA chain terminal methyl groups, CYP4A/4F subfamily members eliminate potentially toxic, excess non-esterified FFAs that could disrupt mitochondrial function and inhibit ATP synthesis (Sanders et al., 2006; Weinberg, 2006; Hsu et al., 2007). Thus, the increase in *CPY4A11* protein expression in the LP group indicates that the ability of liver cells to oxidize non-esterified FAs was enhanced. ω-Oxidation is increased in non-alcoholic fatty liver disease (Kohjima et al., 2007). In the LP group, increased FA oxidation may induce lipid accumulation in the liver and suppress the synthesis of milk fat, although the specific mechanisms remain to be determined. 20-Hydroxyeicosatetraenoic acid (20-HETE) is a product of arachidonic acid that is hydroxylated by *CYP4A11* catalysis, which can stimulate the production of superoxide and inflammatory cytokines, inhibit endogenous nitric oxide synthase, and promote oxidative stress (Lasker et al., 2000; Singh et al., 2007; Cheng et al., 2008; Ishizuka et al., 2008). 20-HETE metabolites can enhance the hypertrophy of mature inflamed adipocytes in populations of mesenchymal stem cells undergoing adipogenic differentiation through pro-adipogenic effects (Kim et al., 2013). 20-HETE and its metabolites also activate PPARγ (Ashley et al., 2002; Fang et al., 2007) to induce adipogenesis. The up-regulation of FAM120A protein in the LP group suggests that 20-HETE plays a role in regulating the metabolism of lipid precursors in the liver, with the elevation in the *CYP4A11* protein level reflecting activation of downstream signaling. The activation of *PLCB2* is thought to play an important role in the regulation of glucose-induced insulin secretion (Zawalich et al., 1997). The observed up-regulation of *PLCB2* in the HP group may be responsible for the down-regulation of *PCK1* and *CYP4A11*, which are inhibited by increased insulin secretion in the liver (Lobato et al., 1985; Gainer et al., 2005).

Integrated Analysis of DEGs and DEPs

Only *SLC22A1* and *HSPA13* were identified in DEGs and DEPs, and their trends were consistent: *SLC22A1* was significantly down-regulated in the HP group, while *HSPA13* was significantly up-regulated in the HP group, both at the transcriptional and proteomic levels. *SLC22A1* is one of three similar cation transporter genes located in a cluster on chromosome 9. Polyspecific organic cation transporters in the liver, kidney, intestine, and other organs are critical for the elimination of endogenous small organic cations. The encoded protein contains 12 putative transmembrane domains and is a plasma integral membrane protein. A previous study showed that loss of OCT1 (*SLC22A1*) caused an increase in the ratio of AMP to ATP, activated the energy sensor AMP-activated kinase (AMPK), and substantially reduced triglyceride (TG) levels in livers from healthy mice (Ligong et al., 2014). An important paralog of this gene is *SLC22A7*, which is involved in milk fat synthesis in the liver (Liang et al., 2017). *HSPA13*, A member of the HSP 70 family, is mainly involved in stress-induced protective responses (Yan et al., 2015). Migdalska et al. (2012) found in the a mouse partial monosomy model for human chromosome 21q11.2-q21.1 that down-regulated expression of HSPA13 resulted in severe liver fatty changes and thickened subcutaneous fat when mice were fed a high-fat diet (HFD), then suspected that this gene may regulate fat deposition. This is consistent with our study finding that the expression level of *HSPA13* in LP group was significantly lower than that in HP group, both at the transcription level and protein level. In addition, HSPA13 is localized in microsomes (Kampinga et al., 2009) and may be one of the upstream factors for further inducing ω-oxidation in LP group, but its role in CPY4A11, FAM120A, and other ω-oxidation-related proteins in this study needs to be further studied. We found that 53 genes were found in DEGs and non-significantly different proteins, and 81 genes were found in non-significantly different genes and DEPs, which indicated the imbalance between proteomic and transcriptome data. There are two possible reasons: First, post-translational modifications (PTMs) play a vital role in the structure, activity and function of protein, and the activity of protein suggests that it is not the level of its expression that determines its function. Second, the currently multi-omics technologies still need to be improved, which not only requires more accurate quantitative technologies but also needs to detect the PTM level of proteins, which is the focus of our next research.

CONCLUSIONS

In this study, we identified genes and proteins involved in the regulation of milk composition and production in Chinese Holstein cows. Through RNA-seq and TMT analyses of liver tissue we generated transcriptomic and proteomic profiles that revealed the regulatory relationships between DEGs and DEPs as well as several key candidate regulatory molecules (*SLC22A1*, *MAPK9*, *PPARGC1A*, *FOXO1*, *SOCS1*, *SOCS2*, *CREB1*, *HNF4A*, *HNF4G*, *GADD45A*, *DUSP1*, *PDGF*, *SYBU*, *DDIT4*, *BAMBI*, *MTHFR*, *SLC27A2*, *PCK1*, *CPT2*, *SIRT3*, *CYP4A11*, and *PLCB2*) and pathways (insulin, insulin resistance, PI3K/Akt, MAPK, prolactin, mTOR, and PPAR) associated with milk fat synthesis. These results can serve as a basis for breeding Holstein cows that produce milk with abundant essential fats, proteins, and other nutrients.

ETHICS STATEMENT

All procedures involving the handling of experimental animals were conducted in accordance with and were approved by the Animal Welfare Committee of China Agricultural University (permit no. DK996).

AUTHOR CONTRIBUTIONS

SZ conceived and designed the study and revised the manuscript. CZ and DS performed the phenotype collection, sample collection, and data analysis and drafted the manuscript. CL participated in

the experimental design and drafted the manuscript. SL, WC, MC, and SS participated in sample collection. All authors have read and approved the final manuscript.

FUNDING

This work is supported by the 863 project (2013AA102504), the National Science and Technology Programs of China (2011BAD28B02), National Key Technologies R & D Program (2012BAD12B01), Beijing Dairy Industry Innovation Team, China Agricultural Research System (CARS-36), and Xinjiang Province Key Technology Integration and Demonstration Program (201230116). We are deeply grateful to all donors who participated in this program.

ACKNOWLEDGMENTS

The authors thank their colleagues in the molecular quantitative genetics team at the National Engineering Laboratory for Animal Breeding of China Agricultural University and all contributors of the present study. We thank for technical support in mass spectroscopy from Shanghai Bioprofile Technology Company Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00672/full#supplementary-material

REFERENCES

Anderson, S. M., Rudolph, M. C., Mcmanaman, J. L., and Neville, M. C. (2007). Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! *Breast Cancer Res.* 9, 204. doi: 10.1186/bcr1653

Aoyama, T., Tsushima, K., Souri, M., Kamijo, T., Suzuki, Y., Shimozawa, N., et al. (1994). Molecular cloning and functional expression of a human peroxisomal acyl-coenzyme A oxidase. *Biochem. Biophys. Res. Commun.* 198, 1113–1118. doi: 10.1006/bbrc.1994.1158

Arun, S. J., Thomson, P. C., Sheehy, P. A., Khatkar, M. S., Raadsma, H. W., and Williamson, P. (2015). Targeted analysis reveals an important role of JAK-STAT-SOCS genes for milk production traits in Australian dairy cattle. *Front. Genet.* 6, 342. doi: 10.3389/fgene.2015.00342

Ashley, L., Shouzuo, W., Mei-Hui, H., Johnson, E. F., Krishna, M. U., Falck, J. R., et al. (2002). The CYP4A isoforms hydroxylate epoxyeicosatrienoic acids to form high affinity peroxisome proliferator-activated receptor ligands. *J. Biol. Chem.* 277, 35105–35112. doi: 10.1074/jbc.M201575200

Beale, E. G., Harvey, B. J., and Forest, C. (2007). PCK1 and PCK2 as candidate diabetes and obesity genes. *Cell Biochem. Biophys.* 48, 89–95. doi: 10.1007/s12013-007-0025-6

Belury, M. A. (2002). Dietary conjugated linoleic acid in health: physiological effects and mechanisms of action. *Annu. Rev. Nutr.* 22, 505–531. doi: 10.1146/annurev.nutr.22.021302.121842

Bionaz, M., and Loor, J. J. (2011). Gene networks driving bovine mammary protein synthesis during the lactation cycle. *Bioinform. Biol. Insights* 5, 83–98. doi: 10.4137/BBI.S7003

Blott, S., Kim, J.-J., Moisio, S., Schmidt-Küntzel, A., Cornet, A., Berzi, P., et al. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163, 253–266.

Cecchinato, A., Macciotta, N., Mele, M., Tagliapietra, F., Schiavon, S., Bittante, G., et al. (2019). Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. *J. Dairy Sci.* 102 (6), 5254–5265. doi: 10.3168/jds.2018-15867

Cheng, J., Ou, J.-S., Singh, H., Falck, J. R., Narsimhaswamy, D., Pritchard, K. A., Jr., et al. (2008). 20-hydroxyeicosatetraenoic acid causes endothelial dysfunction *via* eNOS uncoupling. *Am. J. Physiol. Heart Circ. Physiol.* 294, H1018–H1026. doi: 10.1152/ajpheart.01172.2007

Cho, E., Lee, K., Choi, J., Jeon, H., Lee, S., Cho, Y., et al. (2015). Novel SNPs in the growth arrest and DNA damage-inducible protein 45 alpha gene (GADD45A) associated with meat quality traits in Berkshire pigs. *Genet. Mol. Res.* 14, 8581–8588. doi: 10.4238/2015.July.31.6

Conte, G., Mele, M., Chessa, S., Castiglioni, B., Serra, A., Pagnacco, G., et al. (2010). Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian Brown cattle. *J. Dairy Sci.* 93, 753–763. doi: 10.3168/jds.2009-2581

Crisà, A., Ferrè, F., Chillemi, G., and Moioli, B. (2016). RNA-sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC Vet. Res.* 12, 264. doi: 10.1186/s12917-016-0881-7

De Brachene, A. C., Dif, N., De Rocca Serra, A., Bonnineau, C., Velghe, A. I., Larondelle, Y., et al. (2017). PDGF-induced fibroblast growth requires monounsaturated fatty acid production by stearoyl-CoA desaturase. *FEBS Open Bio* 7, 414–423. doi: 10.1002/2211-5463.12194

Dorland, H. A., Van, Richter, S., Morel, I., Doherr, M. G., Castro, N., and Bruckmaier, R. M. (2009). Variation in hepatic regulation of metabolism during the dry period and in early lactation in dairy cows. *J. Dairy Sci.* 92, 1924–1940. doi: 10.3168/jds.2008-1454

Fang, X., Dillon, J. S., Hu, S., Harmon, S. D., Yao, J., Anjaiah, S., et al. (2007). 20-Carboxy-arachidonic acid is a dual activator of peroxisome proliferator-activated receptors α and γ. *Prostaglandins Other Lipid Mediat.* 82, 175–184. doi: 10.1016/j.prostaglandins.2006.05.002

Gainer, J. V., Bellamine, A., Dawson, E. P., Womble, K. E., Grant, S. W., Wang, Y., et al. (2005). Functional variant of CYP4A11 20-hydroxyeicosatetraenoic acid synthase is associated with essential hypertension. *Circulation* 111, 63–69. doi: 10.1161/01.CIR.0000151309.82473.59

Gao, Y. H., Jiang, J. P., Yang, S. H., Hou, Y. L., Liu, G. E., Zhang, S. G., et al. (2017). CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC Genomics* 18, 265–277. doi: 10.1186/s12864-017-3636-3

Graber, M., Kohler, S., Kaufmann, T., Doherr, M. G., Bruckmaier, R. M., Dorland, H. A., et al. (2010). A field study on characteristics and diversity of gene expression in the liver of dairy cows during the transition period. *J. Dairy Sci.* 93, 5200–5215. doi: 10.3168/jds.2010-3265

Greenfield, R., Cecava, M., and Donkin, S. (2000). Changes in mRNA expression for gluconeogenic enzymes in liver of dairy cattle during the transition to lactation1. *J. Dairy Sci.* 83, 1228–1236. doi: 10.3168/jds.S0022-0302(00)74989-7

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J.-J., Kvasz, A., et al. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc. Natl. Acad. Sci.* 101, 2398–2403. doi: 10.1073/pnas.0308518100

Hammon, H., Stürmer, G., Schneider, F., Tuchscherer, A., Blum, H., Engelhard, T., et al. (2009). Performance and metabolic and endocrine changes with emphasis on glucose metabolism in high-yielding dairy cows with high and low fat content in liver after calving. *J. Dairy Sci.* 92, 1554–1566. doi: 10.3168/jds.2008-1634

Hanson, R. W., and Reshef, L. (1997). Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. *Annu. Rev. Biochem.* 66, 581–611. doi: 10.1146/annurev.biochem.66.1.581

Hardwick, J. P. (2008). Cytochrome P450 omega hydroxylase (CYP4) function in fatty acid metabolism and metabolic diseases. *Biochem. Pharmacol.* 75, 2263–2275. doi: 10.1016/j.bcp.2008.03.004

He, J., Chen, D. L., Samocha-Bonet, D., Gillinder, K. R., Barclay, J. L., Magor, G. W., et al. (2016). Fibroblast growth factor-1 (FGF-1) promotes adipogenesis by downregulation of carboxypeptidase A4 (CPA4)—a negative regulator of adipogenesis implicated in the modulation of local and systemic insulin sensitivity. *Growth Factors* 34, 210–216. doi: 10.1080/08977194.2017.1285764

Hosseini, A., Tariq, M. R., Da Rosa, F. T., Kesser, J., Iqbal, Z., Mora, O., et al. (2015). Insulin sensitivity in adipose and skeletal muscle tissue of dairy cows in response to dietary energy level and 2, 4-thiazolidinedione (TZD). *PLoS One* 10, e0142633. doi: 10.1371/journal.pone.0142633

Hsu, M.-H., Savas, Griffin, K. J., and Johnson, E. F. (2007). Human cytochrome p450 family 4 enzymes: function, genetic variation and regulation. *Drug Metab. Rev.* 39, 515–538. doi: 10.1080/03602530701468573

Ikoma-Seki, K., Nakamura, K., Morishita, S., Ono, T., Sugiyama, K., Nishino, H., et al. (2015). Role of LRP1 and ERK and cAMP signaling pathways in lactoferrin-induced lipolysis in mature rat adipocytes. *PLoS One* 10, e0141378. doi: 10.1371/journal.pone.0141378

Isackson, P. J., Sutton, K. A., Hostetler, K. Y., and Vladutiu, G. D. (2013). Novel mutations in the gene encoding very long-chain acyl-CoA dehydrogenase identified in patients with partial carnitine palmitoyltransferase ii deficiency. *Muscle Nerve* 47, 224–229. doi: 10.1002/mus.23498

Ishizuka, T., Cheng, J., Singh, H., Vitto, M. D., Manthati, V. L., Falck, J. R., et al. (2008). 20-Hydroxyeicosatetraenoic acid stimulates nuclear factor-κB activation and the production of inflammatory cytokines in human endothelial cells. *J. Pharmacol. Exp. Ther.* 324, 103–110. doi: 10.1124/jpet.107.130336

Jacometo, C., Zhou, Z., Luchini, D., Trevisi, E., Correa, M., and Loor, J. (2016). Maternal rumen-protected methionine supplementation and its effect on blood and liver biomarkers of energy metabolism, inflammation, and oxidative stress in neonatal Holstein calves. *J. Dairy Sci.* 99, 6753–6763. doi: 10.3168/jds.2016-11018

Kampinga, H. H., Hageman, J., Vos, M. J., Kubota, H., Tanguay, R. M., Bruford, E. A., et al. (2009). Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress Chaperones* 14, 105–111. doi: 10.1007/s12192-008-0068-7

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Karcher, E., Pickett, M., Varga, G., and Donkin, S. (2007). Effect of dietary carbohydrate and monensin on expression of gluconeogenic enzymes in liver of transition dairy cows. *J. Anim. Sci.* 85, 690–699. doi: 10.2527/jas.2006-369

Kim, D. H., Puri, N., Sodhi, K., Falck, J. R., Abraham, N. G., Shapiro, J., et al. (2013). Cyclooxygenase-2 dependent metabolism of 20-HETE increases adiposity and adipocyte enlargement in mesenchymal stem cell derived adipocytes. *J. Lipid Res.* 54 (3), 786–793. doi: 10.1194/jlr.M033894

Kinsella, J. (1972). Stearyl CoA as a precursor of oleic acid and glycerolipids in mammary microsomes from lactating bovine: possible regulatory step in milk triglyceride synthesis. *Lipids* 7, 349–355. doi: 10.1007/BF02532654

Kohjima, M., Enjoji, M., Higuchi, N., Kato, M., Kotoh, K., Yoshimoto, T., et al. (2007). Re-evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. *Int. J. Mol. Med.* 20, 351–358. doi: 10.3892/ijmm.20.3.351

Kong, Q., Zhang, H., Zhao, T., Zhang, W., Yan, M., Dong, X., et al. (2016). Tangshen formula attenuates hepatic steatosis by inhibiting hepatic lipogenesis and augmenting fatty acid oxidation in db/db mice. *Int. J. Mol. Med.* 38, 1715–1726. doi: 10.3892/ijmm.2016.2799

Krammer, J., Digel, M., Ehehalt, F., Stremmel, W., Füllekrug, J., and Ehehalt, R. (2011). Overexpression of CD36 and acyl-CoA synthetases FATP2, FATP4 and ACSL1 increases fatty acid uptake in human hepatoma cells. *Int. J. Med. Sci.* 8, 599. doi: 10.7150/ijms.8.599

Lasker, J. M., Chen, W. B., Wolf, I., Bloswick, B. P., Wilson, P. D., and Powell, P. K. (2000). Formation of 20-hydroxyeicosatetraenoic acid, a vasoactive and natriuretic eicosanoid, in human kidney role of cyp4F2 and cyp4A11. *J. Biol. Chem.* 275, 4118–4126. doi: 10.1074/jbc.275.6.4118

Leclerc, D., Christensen, K. E., Cauvi, O., Yang, E., Fournelle, F., Bahous, R. H., et al. (2018). Mild methylenetetrahydrofolate reductase deficiency alters inflammatory and lipid pathways in liver. *Mol. Nutr. Food Res.* 63 (3), e1801001. doi: 10.1002/mnfr.201801001

Lee, S., and Dong, H. H. (2017). FoxO integration of insulin signaling with glucose and lipid metabolism. *J. Endocrinol.* 233, R67–R79. doi: 10.1530/JOE-17-0002

Li, L., He, M. L., Wang, K., and Zhang, Y. S. (2018). Buffering agent *via* insulin-mediated activation of PI3K/AKT signaling pathway to regulate lipid metabolism in lactating goats. *Physiol. Res.* 67, 753–764. doi: 10.33549/physiolres.933715

Liang, R., Han, B., Li, Q., Yuan, Y., Li, J., and Sun, D. (2017). Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Sci. Rep.* 7, 6396. doi: 10.1038/s41598-017-06634-w

Ligong, C., Yan, S., Xiaomin, L., Chen, E. C., Sook Wah, Y., Zur, A. A., et al. (2014). OCT1 is a high-capacity thiamine transporter that regulates hepatic steatosis and is a target of metformin. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9983–9988. doi: 10.1073/pnas.1314939111

Liu, L. L., Lin, Y., Liu, L. X., Wang, L. N., Bian, Y. J., Gao, X. J., et al. (2016). Regulation of peroxisome proliferator-activated receptor gamma on milk fat synthesis in dairy cow mammary epithelial cells. *In Vitro Cell. Dev. Biol. Anim.* 52, 1044–1059. doi: 10.1007/s11626-016-0059-4

Lobato, M. F., Careche, M., Ros, M., Moreno, F. J., and García-Ruíz, J. P. (1985). Effect of prolactin and glucocorticoids on P-enolpyruvate carboxykinase activity in liver and mammary gland from diabetic and lactating rats. *Mol. Cell. Biochem.* 67, 19–23. doi: 10.1007/BF00220981

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Mcgregor, R. A., and Poppitt, S. D. (2013). Milk protein for improved metabolic health: a review of the evidence. *Nutr. Metab.* 10, 46–46. doi: 10.1186/1743-7075-10-46

Mele, M., Conte, G., Castiglioni, B., Chessa, S., Macciotta, N. P. P., Serra, A., et al. (2007). Stearoyl-coenzyme A desaturase gene polymorphism and milk fatty acid composition in Italian Holsteins. *J. Dairy Sci.* 90, 4458–4465. doi: 10.3168/jds.2006-617

Mensink, R. P., Zock, P. L., Kester, A. D., and Katan, M. B. (2003). Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am. J. Clin. Nutr.* 77, 1146–1155. doi: 10.1093/ajcn/77.5.1146

Migdalska, A. M., van der Weyden, L., Ismail, O., White, J. K., Sanchez-Andrade, G., Logan, D. W., et al. (2012). Modeling partial monosomy for human chromosome 21q11.2-q21.1 reveals haploinsufficient genes influencing behavior and fat deposition. *PLoS One* 7, e29681. doi: 10.1371/journal.pone.0029681

Moyes, K. M., Drackley, J. K., Morin, D. E., Rodriguez-Zas, S. L., Everts, R. E., Lewin, H. A., et al. (2010). Mammary gene expression profiles during an intramammary challenge reveal potential mechanisms linking negative energy balance with impaired immune response. *Physiol. Genomics* 41, 161–170. doi: 10.1152/physiolgenomics.00197.2009

Mucunguzi, O., Melouane, A., Ghanemi, A., Yoshioka, M., Boivin, A., Calvo, E.-L., et al. (2017). Identification of the principal transcriptional regulators for low-fat and high-fat meal responsive genes in small intestine. *Nutr. Metab.* 14, 66. doi: 10.1186/s12986-017-0221-3

Nukitrangsan, N., Okabe, T., Toda, T., Inafuku, M., Iwasaki, H., Yanagita, T., et al. (2011). Effect of *Peucedanum japonicum* Thunb on the expression of obesity-related genes in mice on a high-fat diet. *J. Oleo Sci.* 60, 527–536. doi: 10.5650/jos.60.527

Ramayo-Caldas, Y., Fortes, M., Hudson, N., Porto-Neto, L., Bolormaa, S., Barendse, W., et al. (2014). A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. *J. Anim. Sci.* 92, 2832–2845. doi: 10.2527/jas.2013-7484

Sanchez, J., Priego, T., Pico, C., Ahrens, W., De Henauw, S., Fraterman, A., et al. (2012). Blood cells as a source of transcriptional biomarkers of childhood obesity and its related metabolic alterations: results of the IDEFICS study. *J. Clin. Endocrinol. Metab.* 97, E648–E652. doi: 10.1210/jc.2011-2209

Sanders, R.-J., Ofman, R., Duran, M., Kemp, S., and Wanders, R. J. (2006). ω-Oxidation of very long-chain fatty acids in human liver microsomes. Implications for X-linked adrenoleukodystrophy. *J. Biol. Chem.* 281, 13180–13187. doi: 10.1074/jbc.M513481200

Schennink, A., Bovenhuis, H., Leon-Kloosterziel, K. M., Van Arendonk, J. A. M., and Visker, M. H. P. W. (2009a). Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *Anim. Genet.* 40, 909–916. doi: 10.1111/j.1365-2052.2009.01940.x

Schennink, A., Heck, J. M., Bovenhuis, H., Visker, M. H., Van Valenberg, H. J., and Van Arendonk, J. A. (2008). Milk fatty acid unsaturation: genetic parameters and effects of stearoyl-CoA desaturase (SCD1) and acyl CoA: diacylglycerol acyltransferase 1 (DGAT1). *J. Dairy Sci.* 91, 2135–2143. doi: 10.3168/jds.2007-0825

Schennink, A., Stoop, W., Visker, M., Van Der Poel, J., Bovenhuis, H., and Van Arendonk, J. (2009b). Genome-wide scan for bovine milk-fat composition. II.

Quantitative trait loci for long-chain fatty acids. *J. Dairy Sci.* 92, 4676–4682. doi: 10.3168/jds.2008-1965

Schlegel, G., Ringseis, R., Keller, J., Schwarz, F. J., and Eder, K. (2012). Changes in the expression of hepatic genes involved in cholesterol homeostasis in dairy cows in the transition period and at different stages of lactation. *J. Dairy Sci.* 95, 3826–3836. doi: 10.3168/jds.2011-5221

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shirpoor, A., Heshmati, E., Kheradmand, F., Gharalari, F. H., Chodari, L., Naderi, R., et al. (2018). Increased hepatic FAT/CD36, PTP1B and decreased HNF4A expression contributes to dyslipidemia associated with ethanol–induced liver dysfunction: rescue effect of ginger extract. *Biomed. Pharmacother.* 105, 144–150. doi: 10.1016/j.biopha.2018.05.121

Singh, H., Cheng, J., Deng, H., Kemp, R., Ishizuka, T., Nasjletti, A., et al. (2007). Vascular cytochrome P450 4A expression and 20-hydroxyeicosatetraenoic acid synthesis contribute to endothelial dysfunction in androgen-induced hypertension. *Hypertension* 50, 123–129. doi: 10.1161/HYPERTENSIONAHA.107.089599

Stoop, W. M., Schennink, A., Visker, M. H., Mullaart, E., Van Arendonk, J. A., and Bovenhuis, H. (2009). Genome-wide scan for bovine milk-fat composition. I. Quantitative trait loci for short-and medium-chain fatty acids. *J. Dairy Sci.* 92, 4664–4675. doi: 10.3168/jds.2008-1966

Velez, J., and Donkin, S. (2005). Feed restriction induces pyruvate carboxylase but not phosphoenolpyruvate carboxykinase in dairy cows. *J. Dairy Sci.* 88, 2938–2948. doi: 10.3168/jds.S0022-0302(05)72974-X

Weinberg, J. (2006). Lipotoxicity. *Kidney Int.* 70, 1560–1566. doi: 10.1038/sj.ki.5001834

Yan, Z., Wei, H., Ren, C., Yuan, S., Fu, H., Lv, Y., et al. (2015). Gene expression of Hsps in normal and abnormal embryonic development of mouse hindlimbs. *Hum. Exp. Toxicol.* 34, 563–574. doi: 10.1177/0960327114555927

Yang, S.-Q., Chen, Y.-D., Li, H., Hui, X., and Gao, W.-Y. (2018). Geniposide and gentiopicroside suppress hepatic gluconeogenesis *via* regulation of AKT-FOXO1 pathway. *Arch. Med. Res.* 49, 314–322. doi: 10.1016/j.arcmed.2018.10.005

Ying, Y., Li, L., Cao, W., Yan, D., Zeng, Q., Kong, X., et al. (2012). The microtubule associated protein syntabulin is required for glucose-stimulated and cAMP-potentiated insulin secretion. *FEBS Lett.* 586, 3674–3680. doi: 10.1016/j.febslet.2012.08.025

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14. doi: 10.1186/gb-2010-11-2-r14

Zawalich, W. S., Bonnet-Eymard, M., and Zawalich, K. C. (1997). Signal transduction in pancreatic beta-cells: regulation of insulin secretion by information flow in the phospholipase C/protein kinase C pathway. *Front. Biosci.* 2, d160–d172. doi: 10.2741/A180

Zhang, Q., Koser, S. L., and Donkin, S. S. (2016). Propionate induces the bovine cytosolic phosphoenolpyruvate carboxykinase promoter activity. *J. Dairy Sci.* 99, 6654–6664. doi: 10.3168/jds.2016-11103

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhou, Shen, Li, Cai, Liu, Yin, Shi, Cao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 August 2019
doi: 10.3389/fgene.2019.00756

Comprehensive Profiles of mRNAs and miRNAs Reveal Molecular Characteristics of Multiple Organ Physiologies and Development in Pigs

Muya Chen[1,2†], Yi Long Yao[1,2†], Yalan Yang[1,2], Min Zhu[1,2], Yijie Tang[1,2], Siyuan Liu[1,2], Kui Li[3] and Zhonglin Tang[1,2,3]*

[1] Research Centre for Animal Genome, Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [2] Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [3] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:
Yun Li,
Ocean University of China, China
Yanzhi Jiang,
Sichuan Agricultural University,
China
Bin Chen,
Hunan Agricultural University, China

***Correspondence:**
Zhonglin Tang
tangzhonglin@caas.cn

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 12 March 2019
Accepted: 17 July 2019
Published: 28 August 2019

Citation:
Chen M, Yao YL, Yang Y, Zhu M, Tang Y, Liu S, Li K and Tang Z (2019) Comprehensive Profiles of mRNAs and miRNAs Reveal Molecular Characteristics of Multiple Organ Physiologies and Development in Pigs. Front. Genet. 10:756. doi: 10.3389/fgene.2019.00756

The pig (*Sus scrofa*) is not only an important livestock animal but also widely used as a biomedical model. However, the understanding of the molecular characteristics of organs and of the developmental skeletal muscle of the pig is severely limited. Here, we performed a comprehensive transcriptome profiling of mRNAs and miRNAs across nine tissues and three skeletal muscle developmental stages in the Guizhou miniature pig. The reproductive organs (ovary and testis) had greater transcriptome complexity and activity than other tissues, and the highest transcriptome similarity was between skeletal muscle and heart (R = 0.79). We identified 1,819 mRNAs and 96 miRNAs to be tissue-specific in nine organs. Testis had the largest number of tissue-specific mRNAs (992) and miRNAs (40). Only 15 genes and two miRNAs were specifically expressed in skeletal muscle and fat, respectively. During postnatal skeletal muscle development, the mRNAs associated with focal adhesion, Notch signaling, protein digestion, and absorption pathways were up-regulated from D0 to D30 and then down-regulated from D30 and D240, while genes with opposing expression patterns were significantly enriched in the oxidative phosphorylation and proteasome pathways. The miRNAs mainly regulated genes associated with insulin, Wnt, fatty acid biosynthesis, Notch, MAPK, TGF-beta, insulin secretion, ECM–receptor interaction, focal adhesion, and calcium signaling pathways. We also identified 37 new miRNA–mRNA interaction pairs involved in skeletal muscle development. Overall, our data not only provide a rich resource for understanding pig organ physiology and development but also aid the study of the molecular functions of mRNA and miRNA in mammals.

Keywords: mRNAs, miRNAs, multiple organ, skeletal muscle, pig

INTRODUCTION

Intensive transcriptome sequencing is increasingly being used to study the mechanisms of organ physiology and development, with a goal of understanding the genetic expressions of tissue-specific diseases (Lage et al., 2008; Koh et al., 2014). Since the expressions, and thus the functions, of many genes vary between tissue types, the study of tissue-specific genetic expression describes those transcriptional variations and provides insights into the underlying genetic mechanisms in each specific tissue type. Thus, obviously the construction of comprehensive tissue-specific transcriptome profiles for both humans and model organisms is of great importance. Indeed, tissue-specific expression of both protein-coding and non-coding RNAs in both humans and mice has been well studied (Roux et al., 2012; Szabo et al., 2015; Zeng et al., 2016; Iwakiri et al., 2017) and the results have been used to elucidate organ physiologies and diseases. For example, the identification of tissue-specific genes was used to build a tissue-specific gene database for human cancers (Kim et al., 2018). However, while many studies have accumulated both mouse and human tissue-specific transcriptome data (Zhang et al., 2015a), RNA-seq transcriptome analyses across tissues and developmental stages of other mammals are still relatively scarce.

Wild *Sus scrofa* (pig) was domesticated approximately 9,000 years ago and has become one of humankind's most important livestock animals (Giuffra et al., 2000). Because of its similarity with humans in body size, lifespan, anatomy, and other distinct physiological characteristics, the pig has become a model in many disciplines of biomedical research, such as pharmacology, obesity, oncology, cardiology, and many more (Groenen et al., 2012). While pig mRNA and miRNA profiles have been obtained for several tissues, such as adipose, liver, skeletal muscle (Hou et al., 2012; Tang et al., 2015), and testis (Zhang et al., 2015b), those analyses focused mainly on a single organ or developmental stage. Systematic studies of miRNA and mRNA spatiotemporal expression patterns and their interactions in multiple organs and developmental stages are still needed to further understand their physiological functions and development, which would aid both biomedical research and pig husbandry. Interactions between miRNAs and their target mRNAs play important roles in regulating various biological processes (Bai et al., 2015), so identification of those interactions provides insights into the different mechanisms at work in each tissue type (Neville et al., 2011; Wang et al., 2012b).

In this study, we used high-throughput transcriptome sequencing to comprehensively explore *S. scrofa* mRNA and miRNA profiles in nine different tissues and three developmental stages of skeletal muscles. First, we systematically analyzed expression characteristics of protein coding genes (PCGs) and miRNAs in nine tissues, identifying the tissue-specific and -associated PCGs/miRNAs and then exploring the interactions of miRNAs with their target mRNAs in nine organs. Finally, we detected a set of miRNA–mRNA interaction pairs potentially associated with postnatal skeletal muscle development. Overall, this study provides a comprehensive profile of mRNAs and miRNAs in multiple organs and developmental stages in the pig and provides meaningful insights into tissue-specific metabolic regulation at the RNA level.

MATERIALS AND METHODS

Animals and Organ Collection

In this study, we collected nine tissues (fat, heart, kidney, liver, lung, skeletal muscle, ovary, spleen, and testis) from the Guizhou miniature pig, one of the most primitive pig breeds in China and a source of high quality meat, at 240 days of age and two additional skeletal muscle samples at postnatal 0 and 30 days. These samples were collected from three biological individuals at each postnatal date (0, 30, and 240 days). All samples were rapidly isolated and immediately frozen in liquid nitrogen. All animal procedures were performed according to protocols approved by the Biological Studies Animal Care and Use Committee in Beijing Province, China.

Isolation of Total RNA and Construction of RNA-seq Libraries

We extracted total RNA from various tissues at least three times, mixing RNA samples from each tissue type into one group per type, and small RNA library for each tissue group was produced. Polyacrylamide electrophoresis gel was used to purify the fragments of 18–30 nt, and then these fragments were ligated to adaptors on both 5′ and 3′ ends. After reverse-transcription amplification, the PCR products in length of 90-bp were isolated from 4 % agarose gels, and then sequenced on the Illumina HiSeq 2500 platform. The RNA-seq data for mRNA were deposited in the Gene Expression Omnibus (accession codes GSE73763) as our previous reports (Tang et al., 2017; Liang et al., 2017), and the reliability of transcriptome data has was verified by qRT-PCR in the study by Tang et al. (2017).

RNA-seq Data Analysis

First, using custom scripts, we trimmed adapters from all RNA sequencing data and then mapped the processed reads from each sample to the *S. scrofa* reference genome (v10.2) using TopHat2 (Kim et al., 2013) (v2.0.12) with fr-frststrand and the following parameters: mate-inner-dist 20, mate-std-dev 50, microexon-search segment-length 25, and segment-mismatches 2. Alignment results from each sample were then processed using Cufflinks (v2.2.1) with known annotations for transcript assembly (fr-firststrand and min-frags-per-transfrag 3) and then the consensus transcriptome was merged. Finally, we used HTseq-count (v0.6.1) required strand-specific counting (Anders et al., 2015) to quantify genes and transcripts. We then calculated the reads per kilobase million (RPKM), counted on read pairs in cases of paired ends, for the PCGs.

miRNA-seq Data Analysis

Raw sequencing reads were obtained after removing reads without a 3′ primer or insert tag, reads with polyA or 5′ primer contaminants, and reads shorter than 18 nt or longer than 32 bp.

Then, the clean reads that could be annotated and aligned to rRNAs, snoRNAs, and tRNAs in the Rfam database (http://rfam.xfam.org) (Nawrocki et al., 2015) were discarded. We mapped the remaining reads to the *S. scrofa* reference genome (v10.2) using miRDeep2(v2.0.0.8) software (Friedlander et al., 2012). The sequences of known mature miRNAs and their precursors were downloaded from miRBase (http://www.mirbase.org) (Kozomara and Griffiths-Jones, 2014), and the expression level of each miRNA was normalized using the transcripts per kilobase million (TPM) method.

Gene Expression Analyses

In our analyses, all PCG-miRNAs that had an RPKM or TPM greater than 0.1 in at least one sample were considered to be expressed. In addition, Pearson correlation coefficients were calculated to examine the similarities and correlations of mRNA expression in different samples. In our analyses, universally expressed genes are tissue-conserved expressed genes whose RPKM values are greater than 10 in every tissue. For miRNAs, two criteria were used to define universally expressed miRNAs across different tissues: 1) the TPM value in every tissue was more than 1, and 2) the coefficient of variation across all tissues was less than 0.5.

The tissue-associated genes for any given tissue were identified according to a previous study with the *Z*-score cutoff ≥ 1.5 and RPKM ≥ 1 (Li et al., 2014). We identified tissue-specific genes as having RPKM or TPM ≥ 10, with expression levels in a given tissue being greater than 10-fold higher than the mean expression value of any other tissues.

Co-Expression Network Analysis

We used RNA libraries of nine different tissue types, with all samples collected at postnatal day 240, for network construction. Based on the mRNA expression matrix, we constructed a weighted co-expression network using BioLayout Express (3D) with a Pearson correlation threshold cutoff ≥ 0.90 and a Markov clustering algorithm of 2.2.

miRNA–mRNA Interaction Analyses

miRNA–host gene co-expression pairs were identified based on the following pipeline: 1) the miRNA coordinate overlapped 100% of the protein-coding gene; 2) the host gene and the miRNA are transcribed from the same strand of DNA; 3) the miRNA does not have extra copies in other parts of the genome, since the transcription of each copy of the miRNA gene could be regulated by different mechanisms that would confound the results of our analyses; 4) the intragenic miRNAs and the host genes must be expressed in at least five tissues (TPM or RPKM ≥ 0.1); and 5) significant correlations were identified as $p < 0.05$ and $r > 0.6$ (Kaminski et al., 2013; Lin et al., 2016).

mRNA and miRNA pairs were subjected to Pearson correlation analysis and those pairs with $r < -0.5$ were chosen for further investigation. Then, we used the RNAhybrid (v2.1.2) (Kruger and Rehmsmeier, 2006) and TargetScan algorithms (Riffo-Campos et al., 2016) to detect whether the 3′-untranslated region (3′UTR) of the mRNA in each pair matched the seed region of the corresponding miRNA. The pairs that satisfied those two conditions were used in both KEGG pathway analysis and Gene Ontology (GO) enrichment analysis to further investigate the biological processes and functions associated with those negative correlations. These analyses were based on human annotation using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web server (http://david.abcc.ncifcrf.gov/) with the EASE value set to 0.05 (Huang et al., 2007; Huang Da et al., 2009).

Differential Expression Analysis

We used R (v3.2.0) DESeq2 package (Love et al., 2014) for differential expression analyses of PCGs and miRNAs in skeletal muscle at three developmental (D) stages (D0, D30, and D240 postnatal days). Analysis was performed between muscle_D0 and muscle_D30 and between muscle_D30 and muscle_D240. In this study, significant differentially expressed genes (DEGs) met the criteria log2-FC ≥ 1 and FDR < 0.05, where FC is fold change and FDR is false discovery rate, and those acceptable DEGs were further examined with GO and KEGG analyses, as above.

Vector Construction, Cell Culture, and Dual Luciferase Reporter Assay

Two miRNA–mRNA interaction pairs (*ACTN4*/ssc-miR-133a-3p *and Prox1*/ssc-miR-338) were randomly selected to verify the expression profiles of miRNA and its target mRNA. Through the PCR method, the 3′UTR fragments (3′UTR-wt) flanking miRNA binding sites of these two genes were amplified and then cloned into pmirGLO Dual-Luciferase Vector through the Homologous Recombination Kit (Qingke, China). The mutant types of these two genes with the 3′UTR region (3′UTR-edt) were made by the Homologous Recombination Kit (Qingke, China) and confirmed by sequencing. Primer sequences are listed in **Table 1**.

The HEK293 cells were cultured at 37°C with Dulbecco's modified Eagle's medium (Sigma), 10% FBS (Gibco), 1% penicillin/streptomycin (Gibco), and 5% CO_2. The miR-and -133b mimics (double-stranded RNA oligonucleotides) and negative control duplexes were synthesized by GenePharma. The pmirGLO-3′UTR-wt, pmirGLO-3′UTR-mt and miRNA (mimic/negative control) were co-transfected into HEK293 cells. The co-transfection assays were performed in 12-well plates with Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions and harvested after 24 h. Finally, the dual-luciferase assay system (Promega) was used to examine the activity of renilla and firefly luciferase.

RESULTS

Overview of mRNA and miRNA Profiling

To systematically investigate genome-wide expression profiles of *S. scrofa* PCGs and miRNAs, we first performed RNA-seq and small RNA-seq on nine tissues, as well as on skeletal muscle tissue from three developmental stages of Guizhou miniature pigs. A total of 824,887,548 reads were obtained for RNA-seq analysis,

TABLE 1 | Primer sequences.

Name	Primer sequences (5′–3′)	Application
ACTN4 3′UTR	F: TCTAGTTGTTTAAACGAGCTCGCCTCTTGCTCCCGTAAT R: CAGGTCGACTCTAGACTCGAGGGAGCAAAACCATCCACTA	Amplification and vector construct
PROX1 3′UTR	F: TCTAGTTGTTTAAACGAGCTCGTAGTCGCAGTCCCCTTT R: CAGGTCGACTCTAGACTCGAGTAAACTAAAGGCGGAAGG	Amplification and vector construct
ACTN4 3′UTR MT	F: TCTAGTTGTTTAAACGAGCTCTATGTTCTGAAATCGTAGTT R: CAGGTCGACTCTAGACTCGAGCCCTCCCTTGCGAACAC	Mutation vector construct
PROX1 3′UTR MT	F: TCTAGTTGTTTAAACGAGCTCTGGGTCTCTGAAAGTT R: CAGGTCGACTCTAGACTCGAGCCCCAGAAGGGTGGTTTATC	Mutation vector construct

as we described in our previous study (Liang et al., 2017). We then measured the expression abundance of PCGs in each tissue (**Figure 1A**), detecting 18,576 expressed PCGs (RPKM > 0.1) representing 85.97% of annotated PCGs in the porcine reference genome, and 46.14% of these PCGs were constitutively expressed through all selected tissues. The numbers of expressed PCGs ranged from 10,845 to 15,552 in the nine different tissues (**Table 2**). The smallest numbers of expressed PCGs were in skeletal muscle and the largest numbers were in the reproductive tissues (ovary and testis). Also, the distribution of genes with high RPKM (>10) values was larger in the reproductive tissues than in the other tissues (**Figure 1B**). These findings indicated that transcriptome complexity and activity in the reproductive system were higher than those in other tissues. We next examined the similarities and correlations between tissues based on global expression profiling. Clustering analysis suggested that the highest transcriptome similarity was shown between skeletal muscle and heart (Pearson correlation, $R = 0.79$), whereas the testis and liver showed the lowest expression correlation (Pearson correlation, $R = 0.41$) (**Figure 1C**).

Small RNA analysis revealed 131,789,681 high-quality clean reads, accounting for 93.64% of the total reads. Analysis of the size distribution of all reads showed that the major class of reads peaked at 22–23 nt within most of the libraries. However, in liver and testis, the majority of clean reads were at 28–30 nt, followed by 22 nt, thus implying that Piwi-interacting RNAs were enriched in those two tissues. For all 11 libraries, 99,812,523 (75.75%) of the clean reads were mapped to the porcine reference genome. The composition of each RNA library (**Figure 1D**) shows our findings that 319 known and 442 novel miRNAs (TPM > 0.1) were expressed in all the tissues we investigated. The number of expressed miRNAs in these libraries ranged from 475 to 594, including 206 to 292 novel miRNAs and 269 to 302 known miRNAs, respectively (**Table 3**). The miRNAs with high RPKM

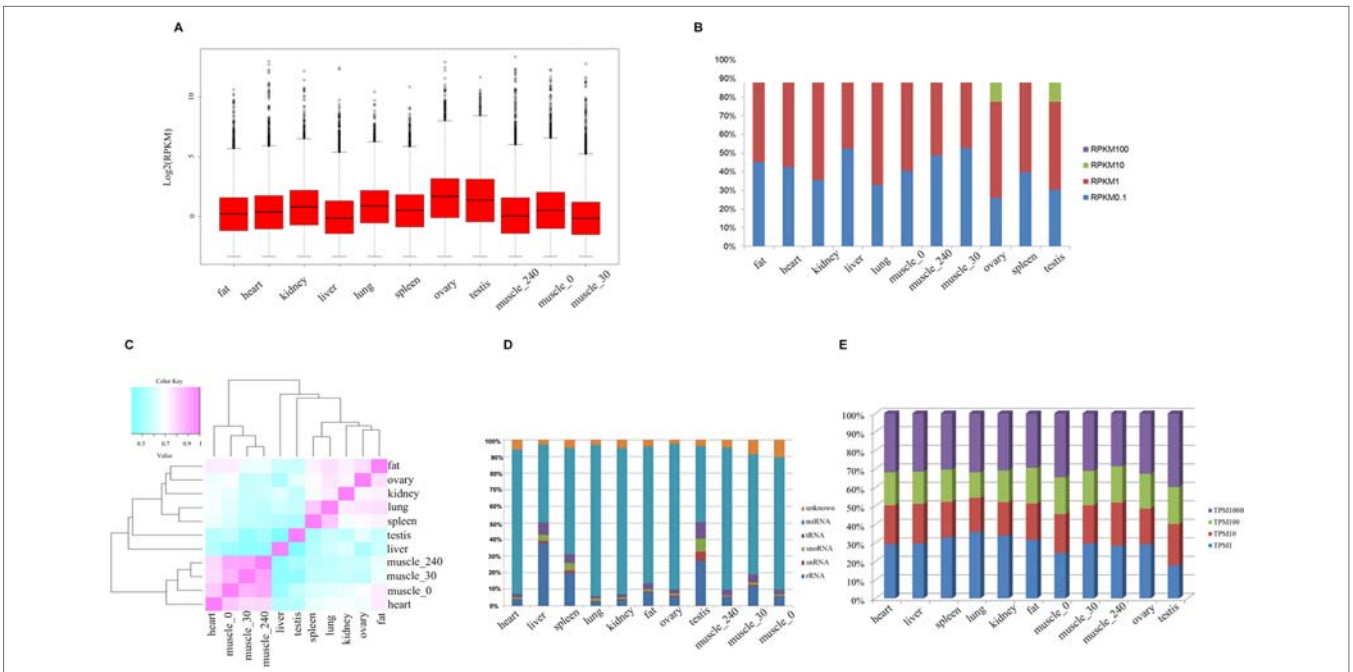


FIGURE 1 | The transcriptome profile of multiple tissues in pigs. **(A)** Genome-wide mean expression value profile. **(B)** Genome-wide RPKM distribution of mRNA. **(C)** Hierarchical clustering generated using Pearson correlation coefficients of log2-transformed RPKM (mRNA) values. **(D)** Composition of the RNA library in each tissue. **(E)** Genome-wide RPKM distribution of miRNA.

TABLE 2 | Numbers of expressed genes in different tissues.

	RPKM1	RPKM10	RPKM100	RPKM1000	Total
Fat	6,015	6,367	843	56	13,281
Heart	5,278	6,032	1,008	89	12,407
Kidney	5,023	7,541	1,436	73	14,073
Liver	6,282	4,867	763	77	11,989
Lung	4,988	8,559	1,474	54	15,075
Muscle_0	4,766	5,746	1,151	124	11,787
Muscle_240	5,297	4,627	820	101	10,845
Muscle_30	5,976	4,752	574	62	11,364
Ovary	3,928	7,792	3,171	260	15,151
Spleen	5,193	6,928	936	55	13,112
Testis	4,722	7,339	3,277	214	15,552

TABLE 3 | Numbers of miRNA identified in different tissues.

Tissue	Novel	Known	Total
Heart	229	283	512
Liver	260	283	543
Spleen	292	302	594
Lung	275	293	568
Kidney	258	290	548
Fat	286	299	585
Muscle_0	216	293	509
Muscle_30	282	294	576
Muscle_240	206	269	475
Ovarium	248	292	540
Testis	222	283	505
Total	439	319	758

(>10) values had a larger distribution in testis than in other tissues (**Figure 1E**).

Universally and Specifically Expressed mRNAs and miRNAs Across Tissues

Focusing on universally expressed mRNAs and miRNAs, we found that 209 mRNAs (**Table S1**), representing tissue-conserved expressed genes whose RPKM values were greater than 10 in all tissues, were abundantly and stably expressed. This dataset included some well-known housekeeping genes such as *GAPDH*, *ACTB*, *RPS18*, *B2M*, *RPL4*, *RPL37*, and *RPL38*. According to GO analysis, these genes were significantly enriched in the areas of translation, peptide biosynthetic, and amide biosynthetic (**Table S2**), indicating that they have important roles in maintaining essential basal cellular functions. Additionally, we identified 43 universally expressed miRNAs (**Table S3**). One of those miRNAs, miR-16, is most likely an important biomarker for several diseases including lung cancer, rheumatoid arthritis, and sepsis (Wang et al., 2012a; Sromek et al., 2017; Dunaeva et al., 2018) in humans; is abundantly expressed in all tissues; and has been used as a control in several systems, including animal models. These miRNAs may serve as candidate reference to normalize miRNA expression across tissues.

In order to capture the functional differences in gene expression between tissues, we next analyzed tissue-specific mRNA and miRNA. As in a previous study (Li et al., 2017), the genes whose abundance in one tissue was more than fourfold the mean expression value of that in other tissues were defined as tissue-specific genes. While for some genes the expression levels were very low, we defined the tissue-specific mRNAs and miRNAs as the genes whose abundance in one tissue was more than 10-fold the mean expression value in other tissues, with RPKM or TPM ≥10. Finally, we identified 1,819 tissue-specific PCGs, with the number ranging from 15 to 992 for a given tissue (**Table S4**). Testis had the largest number of tissue-specific genes (54.5%, 992/1819), compared with other tissues. In contrast, only 15 genes were specifically expressed in skeletal muscle. We found that the expression of *GAPDH* was significantly higher in skeletal muscle than in other tissues, a result consistent with a previous mouse study (Fortes et al., 2016). Meanwhile, we identified 96 tissue-specific miRNAs (including 48 novel and 48 known miRNAs). The number of miRNAs ranged from 2 to 40 for a given tissue. The largest numbers of tissue-specific miRNAs were found in testis and only two tissue-specific miRNAs were found in fat (**Table 4**). The well-known myomiRs (miRNA-1, miR-133a/b, and miR-206) were specifically expressed in skeletal muscle. In testis, the top two tissue-specific miRNAs, miR-34c and miR-202-5p, were shown to possess an important influence in spermatogenesis regulation (Dabaja et al., 2015; Wang et al., 2018b). For fat, only one known miRNA (miR-224), which was reported to have an important role in adipogenesis development, was specifically expressed (Peng et al., 2013).

Tissue-Associated mRNAs Capture the Structure and Functional Features of Different Tissues

A weighted and undirected co-expression network analysis was performed to understand interactions between PCGs, and it generated 229 distinct clusters containing 13,894 nodes (**Figure 2A**). Cluster names were based on the tissues in which the genes were expressed the most. The three largest clusters (including 4,997 nodes) were groups of highly expressed genes in ovary and testis. These results indicated that the majority of PCGs showed a tissue-restricted expression pattern. Thus, we tried to capture the basic characteristics of gene expression for each tissue by using "tissue-associated genes," genes that were highly expressed in one tissue relative to other tissues (see

TABLE 4 | Tissue-specific miRNA in different tissues.

Tissue	Known miRNAs	Novel miRNAs
Heart	ssc-miR-208b, ssc-miR-490, ssc-miR-499-3p, ssc-miR-499-5p, ssc-miR-7136-5p	novel_14_8571, novel_7_25708
Liver	ssc-miR-122, ssc-miR-192, ssc-miR-194b-5p, ssc-miR-365-3p, ssc-miR-885-3p, ssc-miR-885-5p	novel_13_6504, novel_14_8741, novel_9_28064
Spleen	ssc-miR-142-5p, ssc-miR-145-5p, ssc-miR-150, ssc-miR-20b, ssc-miR-342, ssc-miR-7138-5p	novel_10_3375, novel_2_14974, novel_GL896302.1_32316
Lung	ssc-miR-138,ssc-miR-205,ssc-miR-92b-3p	novel_5_19972, novel_5_19974
Kidney	ssc-miR-10a-3p, ssc-miR-10a-5p, ssc-miR-196b-5p, ssc-miR-204, ssc-miR-429	novel_1_778, novel_6_20939
Fat	ssc-miR-224	novel_5_18915
Skeletal muscle	ssc-miR-1, ssc-miR-133a-3p, ssc-miR-133a-5p, ssc-miR-133b, ssc-miR-206	novel_17_12538, novel_2_14142, novel_2_15377, novel_3_17211, novel_X_29629
Ovarium	ssc-miR-132, ssc-miR-135, ssc-miR-212, ssc-miR-7857-3p	novel_8_26925, novel_9_28118
Testis	ssc-miR-137, ssc-miR-153, ssc-miR-182, ssc-miR-202-3p, ssc-miR-202-5p, ssc-miR-216, ssc-miR-217, ssc-miR-221-3p, ssc-miR-222, ssc-miR-34c, ssc-miR-708-3p, ssc-miR-708-5p	novel_11_3521, novel_13_5908, novel_14_7685, novel_15_10051, novel_16_11835, novel_16_11837, novel_2_14220, novel_2_14976, novel_3_16101, novel_4_17487, novel_5_18832, novel_5_20192, novel_5_20217, novel_5_20223, novel_9_28635, novel_9_28958, novel_9_29013, novel_GL893854.2_31046, novel_X_29642, novel_X_29643, novel_X_30198, novel_X_30204, novel_X_30206, novel_X_30208, novel_X_30210, novel_X_30214, novel_X_30215, novel_X_30218

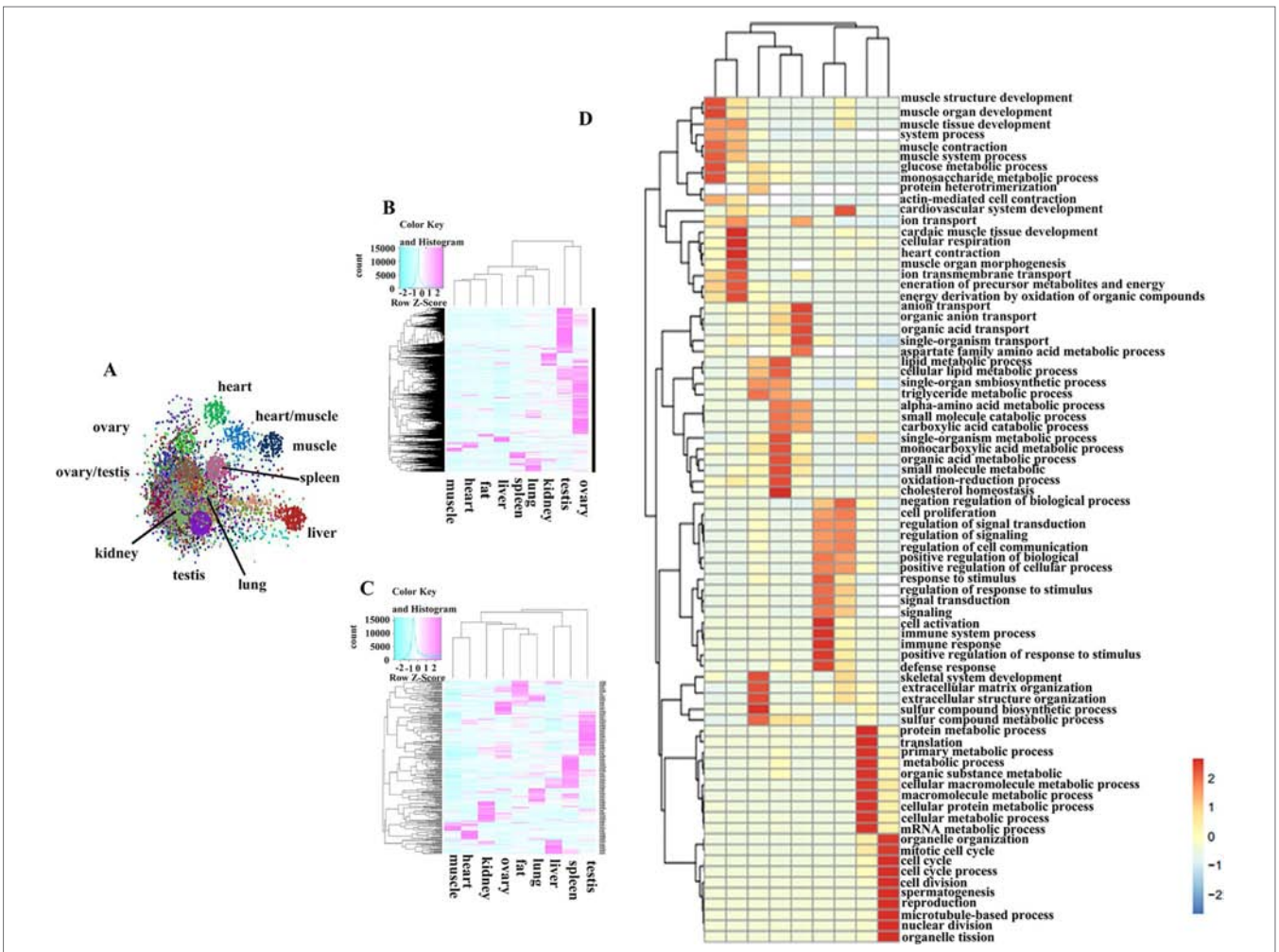


FIGURE 2 | The expression characteristic of tissue-associate genes. **(A)** Co-expression network of the protein coding genes. **(B)** Heatmap of tissue-associated mRNA in different pig tissues. **(C)** Heatmap of tissue-associated miRNA in different pig tissues. **(D)** GO biological process analysis of tissue-associated mRNA.

Materials and Methods). The number of associated PCGs and miRNAs (**Tables S5**, **S6**) for a given tissue ranged from 287 to 5,606 and from 28 to 132, respectively, suggesting that the tissues with higher transcriptional activities, such as testis, had more associated genes (**Figures 2B, C**). GO analysis, based on tissue-associated PCGs, revealed physiological features for each organ (**Figure 2D**). For example, GO terms for muscle tissue development were significantly enriched in skeletal muscle and heart, while the genes associated with spermatogenesis, lipid metabolism, and immune response were enriched in testis, adipose, and spleen, respectively. Also, we observed common GO terms shared in different tissues. For instance, the GO term related to the cell cycle and metabolic processes was obviously enriched in both ovary and testis, confirming that the reproductive tissues were highly proliferative. The GO terms for small-molecule catabolic and other metabolic processes were markedly enriched in liver and kidney. The genes associated with stimulus response and signal transduction were significantly enriched in lung and spleen. In summary, GO terms of tissue-associated genes agreed with the physiologies of the corresponding organs (**Table S7**).

Differentially Expressed mRNA and miRNA During Skeletal Muscle Development

To understand postnatal skeletal muscle development, we assessed the differentially expressed PCGs and miRNAs in skeletal muscle across three developmental stages at 0, 30, and 240 days after birth (D0, D30, and D240, respectively). Between D0 and D30, we detected 1,515 DEGs (**Table S8**), including 911 up-regulated and 604 down-regulated genes, respectively (**Figure 3A**). GO analysis (**Table S9**) suggested that the up-regulated genes were involved mainly in vasculature development, the intracellular signaling cascade, blood vessel development, and enzyme linked receptor protein signaling pathway (**Figure 4A**), and the down-regulated genes were associated mainly with translation, ribonucleoprotein complex biogenesis, and RNA and ncRNA processing (**Figure 4B**). Between D30 and D240, we identified 1,011 DEGs (**Table S10**) including 338 up-regulated and 673 down-regulated genes (**Figure 3B**). According to GO analysis (**Table S9**), up-regulated genes were involved mainly in protein catabolic process, modification-dependent macromolecules, and modification-dependent protein catabolic process (**Figure 4C**). The down-regulated genes were involved mainly in cell

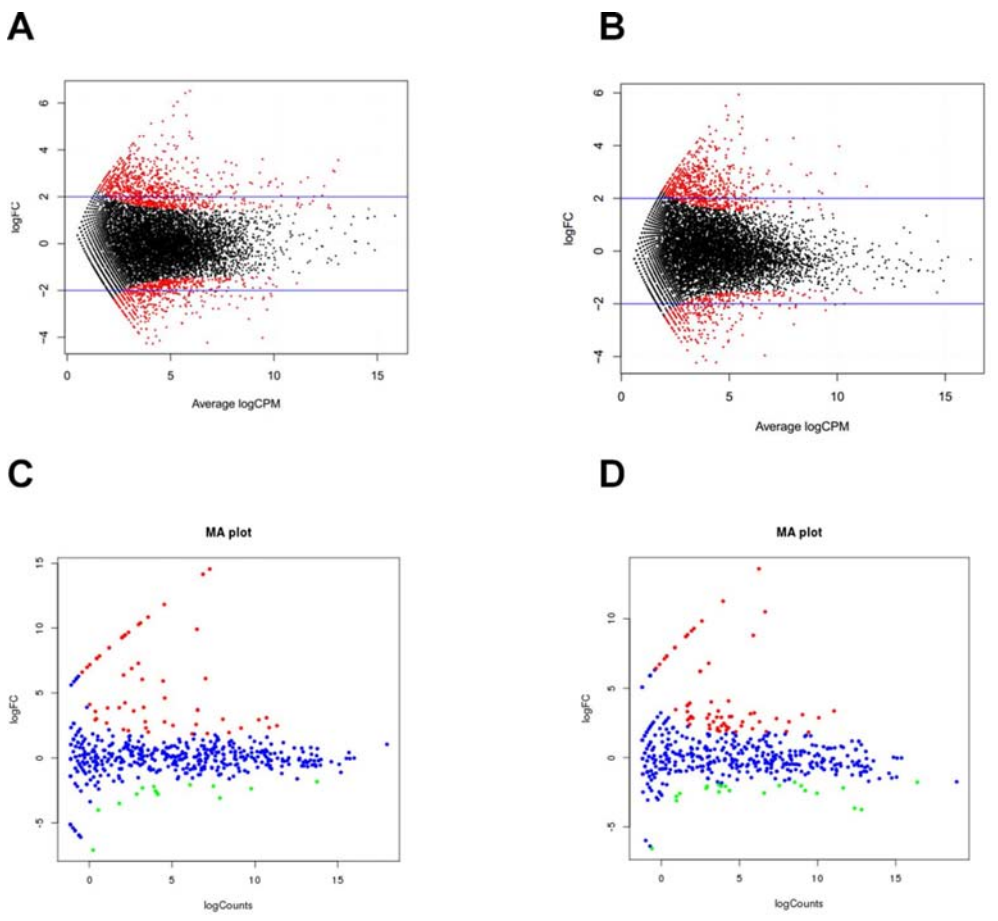


FIGURE 3 | Differentially expressed genes during skeletal muscle development. **(A)** D0 vs. D30 mRNA. **(B)** D30 vs. D240 mRNA. **(C)** D0 vs. D30 miRNA. **(D)** D30 vs. D240 miRNA.

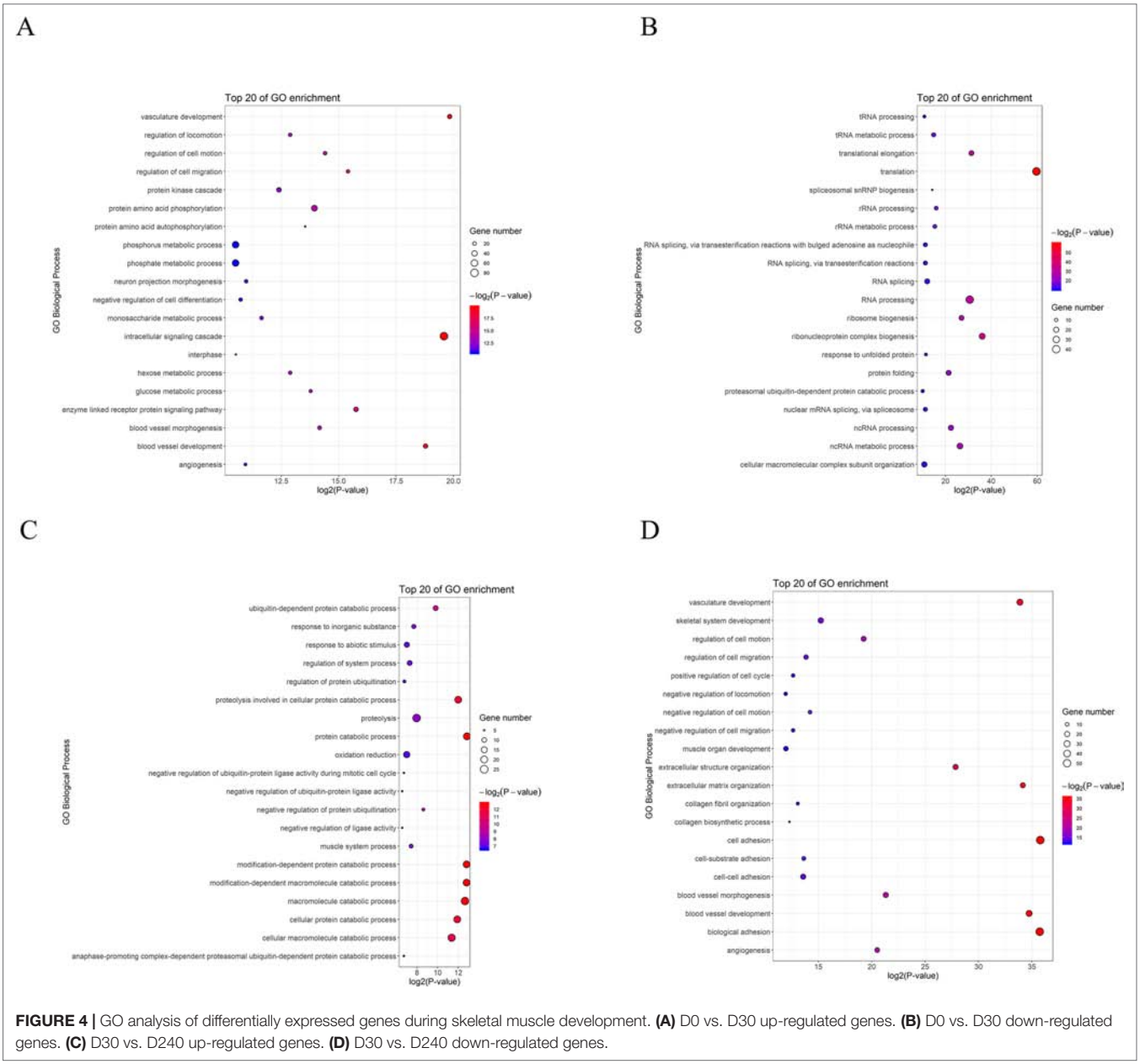


FIGURE 4 | GO analysis of differentially expressed genes during skeletal muscle development. **(A)** D0 vs. D30 up-regulated genes. **(B)** D0 vs. D30 down-regulated genes. **(C)** D30 vs. D240 up-regulated genes. **(D)** D30 vs. D240 down-regulated genes.

adhesion, biological adhesion, and skeletal systems development (**Figure 4D**). These findings indicated that proliferative cell activity decreased, while cellular metabolic ability increased with age during postnatal skeletal muscle development and growth.

Subsequently, we focused on a series of dynamic expression patterns exhibited during skeletal muscle development. Venn diagram (**Figure 5**) analysis for DEGs demonstrated that the greatest overlap (276 genes) occurred between both up-regulated genes in group D0 versus D30 and down-regulated genes in group D30 versus D240 (**Table S11**). The largest cluster of overlapping genes were significantly associated with vasculature, the cardiovascular and circulatory systems, and blood vessel development, and obviously gathered in the focal adhesion, Notch signaling, and protein digestion and absorption pathways (**Tables S12**, **S13**). The second largest overlap cluster contained 94 genes and were present between the down-regulated genes in group D0 versus D30 and the up-regulated genes in group D30 versus D240 (**Table S11**). These overlapping genes functioned in the ATP metabolic process, purine ribonucleoside triphosphate metabolic process, and ribonucleoside triphosphate metabolic process and were significantly enriched in KEGG pathways for oxidative phosphorylation, proteasome, and Parkinson's disease (**Tables S12**, **S13**). However, there were only eight genes up-regulated and seven genes down-regulated throughout D0 to D240 (**Table S11**).

These miRNAs also played an important role in skeletal muscle development. In this study, we identified 70 and 85 differentially expressed miRNAs in groups D0 versus D30 and D30 versus D240, respectively (**Tables S14**, **S15**). Between D0 and D30, 56 miRNAs were up-regulated and 14 were down-regulated (**Figure 3C**). Functional analysis suggested that the genes targeted

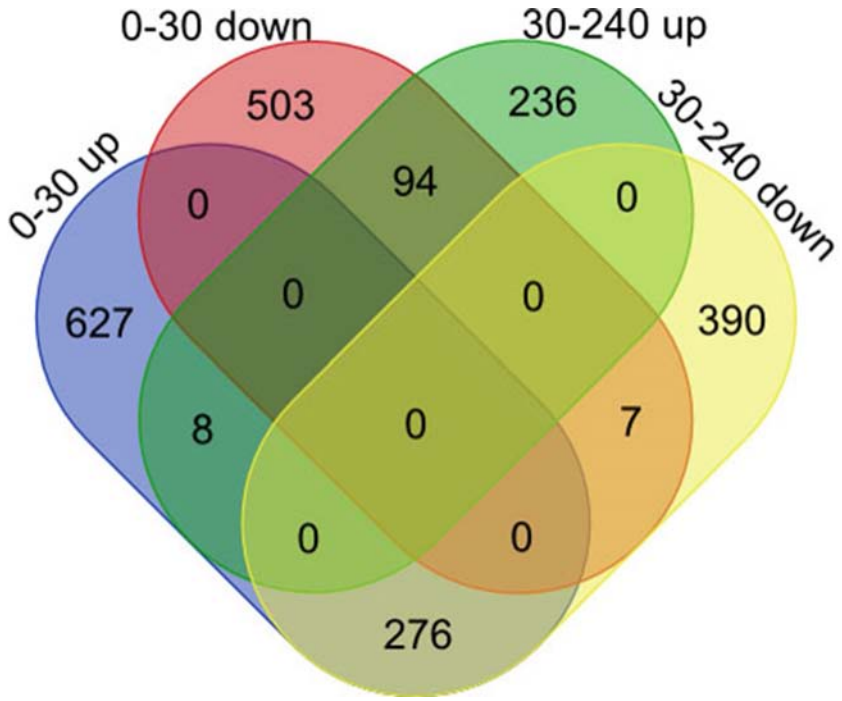


FIGURE 5 | Co-expression pattern analysis for genes during skeletal muscle development.

by down-regulated miRNAs were significantly enriched in the insulin, Wnt, and Notch signaling pathways and in fatty acid biosynthesis, while the target genes for up-regulated miRNAs associated mainly with the MAPK and TGF-beta signaling, insulin secretion, and ECM–receptor interaction pathways. In groups D30 versus D240, we detected 21 up-regulated and 64 down-regulated miRNAs, respectively (**Figure 3D**). The targets for the up-regulated miRNAs were significantly involved in the MAPK, focal adhesion, and insulin signaling pathways, while the targets for the down-regulated miRNAs were enriched mainly in the MAPK, TGF-beta, Notch, calcium, Wnt, insulin secretion, and focal adhesion pathways.

miRNA–mRNA Interaction Network Associated With Skeletal Muscle Development

miRNAs can affect gene expression by inhibiting protein translation or by causing mRNA degradation (Tang et al., 2015). When we evaluated the expression relationships between miRNA and mRNA using Pearson correlations, we detected 253,057 miRNA–mRNA interactions that were negatively correlated ($r < -0.5$) and 2,194 pairs (1,605 mRNAs and 263 miRNAs) with binding sites for miRNAs at mRNA 3′UTRs. GO enrichment analysis suggested that these miRNA–mRNA interactions associated mainly with protein catabolic processes, muscle cell differentiation, and muscle organ development. KEGG analysis revealed that the interactions were significantly enriched in pathways for the citrate cycle, axon guidance, purine and pyruvate metabolisms, as well as the gonadotropin-releasing hormone, MAPK, adipocytokine, and insulin signaling pathways.

In addition, there were 79 miRNA–mRNA interaction pairs that demonstrated functions in muscle cell differentiation and muscle organ development (**Table S16**). Of them, 37 miRNA–mRNA pairs showed significant negative expression correlations ($r < -0.5$) through all three skeletal muscle development stages (**Table 5**). Many mRNA genes have been reported as regulators of either muscle differentiation or development, including *MyoD1* (Blum et al., 2012), *CSRP2* (Herrmann et al., 2006), *MBNL1* (Chen et al., 2016), *FHOD1* (Staus et al., 2011), *MET* (Park et al., 2015), *SOD1* (Sakellariou et al., 2018), *RCAN1* (Emrani et al., 2015), *SGCA* (Fougerousse et al., 1998), *SRF* (Ding et al., 2017), *MEF2A* (Yuan et al., 2014), *MTM1*(Bachmann et al., 2017), *LBX1* (Chao et al., 2011), *MEF2D* (Runfola et al., 2015), *IGFBP5* (Zhang et al., 2017), *PDGFA* (Tallquist et al., 2000), *AMOT* (Wang et al., 2018a), *PDPK1* (Mora et al., 2003), *SIRT2* (Arora and Dey, 2014), *SIX4* (Chakroun et al., 2015), *NOS1* (Villmow et al., 2015), *MYL1* (Burguiere et al., 2011), *ACTA1* (Hu et al., 2015), *PROX1* (Kivela et al., 2016), and *QKI* (Wu et al., 2017). These interaction pairs included 7 known and 27 novel miRNAs, of which the miRNAs ssc-miR-744 (Yang et al., 2015), ssc-miR-497 (Sato et al., 2014), ssc-miR-338 (Mcdaneld et al., 2009), ssc-miR-423-3p (Siengdee et al., 2015), and ssc-miR-133a-3p (Wang et al., 2018c) were reported to have important roles in myogenesis.

Dual Luciferase Reporter Assay Validated the Interaction Between miRNAs and Their Target Genes

We randomly selected two miRNA–mRNA pairs (*ACTN4*/ssc-miR-133a-3p *and Prox1*/ssc-miR-338), which showed significant

TABLE 5 | The expression correlation of miRNA and their target mRNA during three muscle development stages.

mRNA	miRNA	*R* value	mRNA	miRNA	*R* value
MYOD1	novel_6_20992	−0.5390881	MTSS1	novel_9_27936	−0.5860598
HBEGF	novel_6_20992	−0.7102848	SMARCD3	novel_10_2856	−0.5633121
ACHE	novel_X_29788	−0.7475892	RHOQ	novel_10_2856	−0.8857205
SRF	novel_X_29788	−0.9981668	CAPZA2	novel_4_17324	−0.9810114
CALR	novel_15_11149	−0.7613538	LIMK1	novel_9_27762	−0.9571368
GYLTL1B	novel_15_11149	−0.8369125	FHOD1	novel_9_27796	−0.6414053
METTL21A	novel_15_11149	−0.8442253	MEF2A	novel_6_22335	−0.8056134
WASL	novel_15_11149	−0.8928107	MTM1	novel_2_14454	−0.6334166
WASF3	novel_15_11149	−0.8967293	MTM1	novel_2_15515	−0.9702249
RASA1	novel_15_11149	−0.9028681	GPHN	novel_15_11290	−0.7468672
CSRP2	novel_15_11149	−0.9587139	LBX1	novel_16_11622	−0.9452654
EZR	novel_15_11149	−0.9827896	MEF2D	novel_13_6599	−0.9327755
PPP3CB	novel_4_18615	−0.6112719	IGFBP5	novel_4_18039	−0.9461923
MBNL1	novel_4_18615	−0.7985375	PDGFA	novel_13_7087	−0.9090636
FHOD1	novel_12_5322	−0.6604956	PDGFA	novel_5_19528	−0.8911486
ACADM	novel_12_5365	−0.9987412	AMOT	novel_3_15819	−0.577509
MET	novel_GL896243.1_30542	−0.6683036	SPTAN1	novel_2_15576	−0.8980043
SOD1	novel_GL896243.1_30542	−0.9586679	MTSS1	novel_12_5119	−0.7397887
RCAN1	novel_12_5576	−0.9999364	PDPK1	novel_17_12538	−0.9859897
SGCA	novel_1_1367	−0.9808128	WIPF1	novel_14_8940	−0.8267719
SRF	novel_1_1367	−0.8138354	CDC42BPB	novel_17_12654	−0.9731097
SIRT2	novel_X_29783	−0.8866714	SIX4	ssc-miR-744	−0.8646932
RXRG	ssc-miR-296-3p	−0.9848669	PROX1	ssc-miR-338	−0.9395592
NOS1	ssc-miR-296-3p	−0.963125	QKI	ssc-miR-423-3p	−0.9453357
MYL1	ssc-miR-497	−0.7053538	ACTN4	ssc-miR-133a-3p	−0.6850723
ACTA1	ssc-miR-1307	−0.7243173			

negative expression correlations ($r < -0.5$) through all three skeletal muscle developmental stages, to verify whether the interaction between them were really exit. The dual luciferase reporter assay successfully validated the interaction between miRNA and their target genes. As shown in **Figure 6**, by binding to the 3′UTR region, all of the two miRNAs could markedly decrease the luciferase activity of the wild-type target genes (3′UTR-wt), while for the mutant type (3′UTR-mt), this repression was relieved. These results further confirmed the interaction between miRNA–mRNA pairs, which we discovered in the skeletal muscle development.

DISCUSSION

Transcriptome profiling is a good way to understand physiological functions and developmental regulations of organ tissues in plants and animals (Mcloughlin et al., 2014; Santos et al., 2014). The pig is not only an important livestock animal but also an important model organism in biomedical research. Therefore, a comprehensive atlas of gene expression for *S. scrofa* tissues and developmental stages is essential for both breeding and biomedical research (Yang et al., 2016). Here, we used RNA-seq to profile mRNAs and miRNAs found in nine different organ tissues and in three developmental stages of the Guizhou miniature pig, a breed widely used as a model organism in biomedical research. In all, the 18,576 PCGs we detected represent 85.97% of the annotated PCGs in the porcine reference genome. The numbers of PCGs ranged from 10,845 to 15,552 in different tissues and developmental stages, respectively, but only 46.14% of PCGs were constitutively expressed in all nine tissues. This indicates that PCG expressions are greatly temporally and spatially specific. Interestingly, the reproductive organs (ovary and testis) harbored the most complex transcriptomes, a finding consistent with those in humans and rats in which the testis and ovary also expressed the most genes (Yu et al., 2014).

We further analyzed the tissue-associated, tissue-specific, and universally expressed mRNA and miRNA across different tissues and showed that tissue-specific and -associated mRNAs and miRNAs are needed to maintain specific functions in a given tissue type. In this study, the numbers of tissue-specific and tissue-associated PCGs ranged from 15 to 992 and from 287 to 5,606, respectively. High variances reflected the differences in cell homogeneity and activity between different tissues. Tissues with higher transcriptional activities, such as testis and ovary, contained more associated and specific genes than other tissue types. The complex transcript of the Guizhou miniature pig testis was similar to that of other pig varieties. For instance, during the testis development of the Shaziling pig (a Chinese indigenous breed), 8,343 DEGs were identified and more than 50,000 miRNA–mRNA interaction sites were predicted (Ran et al., 2015). Additionally, many of these tissue-associated and tissue-specific genes are well correlated with physiological functions of each organ. For example, the testis-associated gene *PAK2* (p21-activated kinase 2) was reported to play a crucial regulatory role in porcine spermatogenesis apoptosis and when the *PAK2* gene was knocked down by related siRNA, the mitotic activity for Sertoli cells was significantly repressed (Ran et al., 2018b).

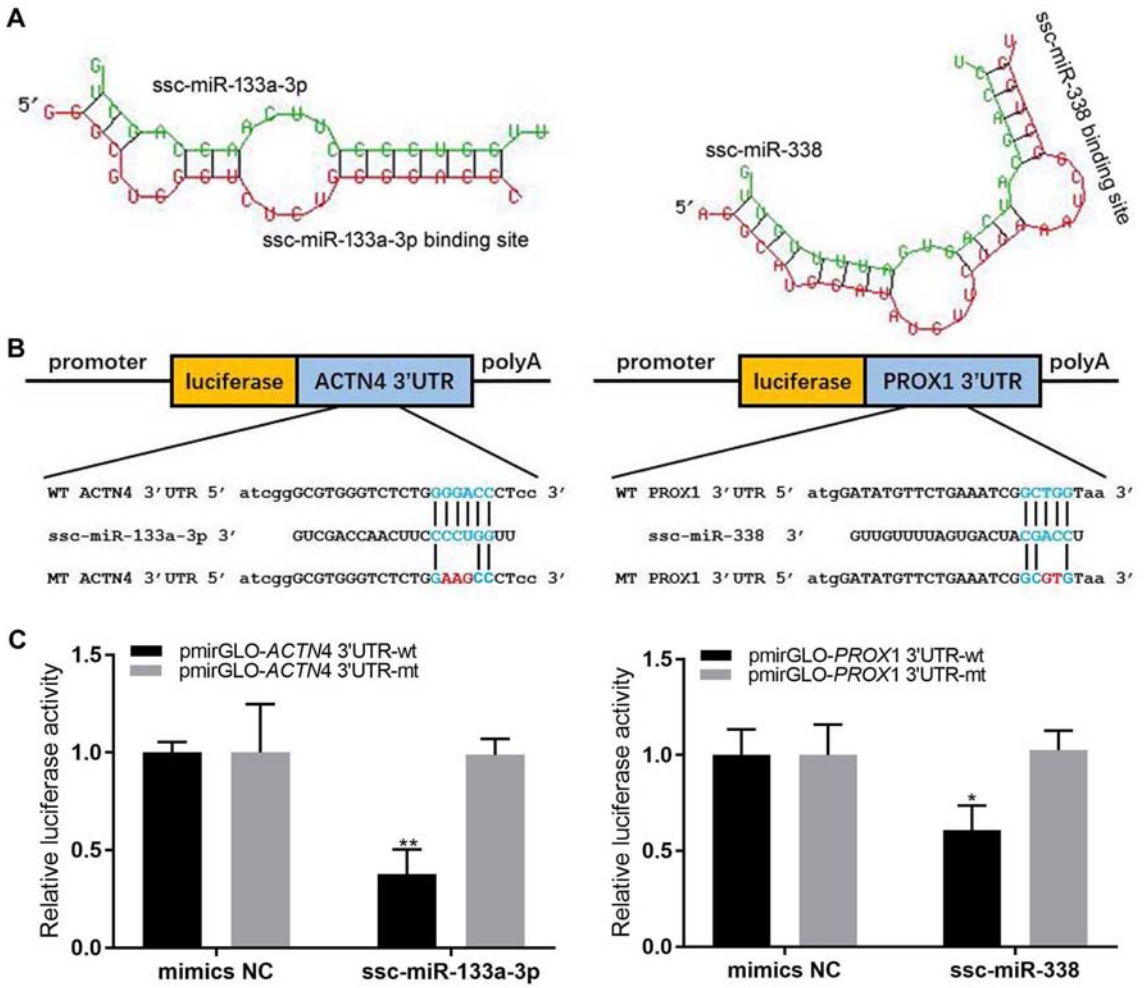


FIGURE 6 | Dual luciferase reporter assay. **(A)** The secondary structure for miRNA and target gene 3′UTR. **(B)** Schematic of the wild-type and mutation-type 3′UTR vector with miRNA binding sites. Red is the mutation sites, and blue is the wild-type binding sites. **(C)** luciferase assays were performed.

Interestingly, while *PAK2* is one target gene of miR-26a, another miR-26a target gene, *ULK2*, was also a testis-associated gene in our study. *ULK2*, when knocked down, will inhibit swine Sertoli cell autophagy (Ran et al., 2018a). Meanwhile the testis-specific genes identified in this study such as *SPEM1*, *TNP1*, *PRM1*, *DAZL* (Hashemi et al., 2018), and *CABYR* (Shen et al., 2019) were also published as specific expression in human or mouse. As we know, tissue-specific functions are a result of specific expression and regulation of genes across an organism's lifespan. The transcriptome's degree of correlation suggests both similar and different biological functions between tissues. These data aid in the understanding of organ physiologies and molecular functions of genes in mammals.

Skeletal muscle is an important organ for maintaining movement and energy metabolism in animals (Liu et al., 2018), and pig skeletal muscle is a protein resource for humans (Tang et al., 2017; Yang et al., 2017). Thus, a systematic study of skeletal muscle development is essential to improving animal breeding as well as aiding biomedical research. Many studies have suggested that the PCGs, miRNAs, and the interactions between them are most important for cellular regulatory processes (Hou et al., 2016). Nielsen et al. (2010) analyzed the miRNA in pig longissimus dorsi by using deep sequencing, and they found that highly expressed miRNAs were involved in skeletal muscle development and regeneration (Nielsen et al., 2010). However, the understanding of skeletal muscle development based on a comprehensive profiling of mRNAs and miRNAs had been largely unclear. To fill that void, we carried out RNA-seq and small RNA-seq analysis on pig skeletal muscle at 0, 30, and 240 days after birth. In the D0 versus D30 and D30 versus D240 groups, 1,515 and 1,011 mRNAs, respectively, were differentially expressed. Functional analysis suggested the presence of significant differences in physiological characteristics at different developmental stages. Between D0 and D30, for example, genes functionally associated with translation, ribonucleoprotein complex biogenesis, and RNA and ncRNA processing were down-regulated, and in the D30 versus D240 groups, down-regulating genes were obviously involved in cell and biological adhesion and in skeletal system development.In the D0 versus D30 and D30 versus D240

groups, we detected 70 and 85 differentially expressed miRNAs, respectively. The miRNAs regulated biological processes by binding mRNAs at the 3′UTR. In the current study, 2194 negatively correlated ($r < -0.5$) miRNA–mRNA interaction pairs with binding sites for miRNAs at mRNA 3′UTRs were predicted. Of those pairs, 37 new miRNA–mRNA interaction pairs were associated with muscle cell differentiation and muscle organ development and were negatively correlated ($r < -0.5$) in the D0, D30, and D240 groups. Most of the predicted target mRNA in these pairs were reported to function in muscle. For instance, *SRF* and *MBNL1* (serum response factor and muscleblind-like splicing regulator 1) genes are reported to regulate muscle atrophy in mice (Collard et al., 2014), *AMOT* (the angiomotin gene) may influence human aortic smooth muscle cell migration (Wang et al., 2018a), and *QKI* (the protein quaking gene) regulates smooth muscle cell differentiation (Wu et al., 2017). Notable miRNAs that we found include ssc-miR-744, which is reported to significantly up-regulate in muscles after ischemia–reperfusion injury (Yang et al., 2015); ssc-miR-195, which induces postnatal quiescence of skeletal muscle stem cells (Sato et al., 2014); and ssc-miR-423-3p and ssc-miR-133a-3p, which each showed high correlations with mouse skeletal muscle C2C12 myoblast differentiation (Siengdee et al., 2015; Wang et al., 2018c). By targeting PCGs, miRNAs play important roles in regulating the complex processes of muscle development. Analysis of miRNA and mRNA expression profiles together was an effective way to minimize false-positive rates in miRNA–mRNA interaction pair predictions. In order to discover more miRNA–mRNA interaction pairs, we first analyzed the transcriptomes of both miRNA and mRNA in nine different tissues and then validated the miRNA–mRNA interaction pairs associated with muscle development in three different muscle development stages. For further verification, we also randomly selected two miRNA–mRNA interaction pairs to verify the expression profiles of each miRNA and its target mRNA by using a dual luciferase reporter assay. The results showed that, by binding to the 3′UTR region, miRNA could markedly decrease the luciferase activity of the target genes. Hence, the miRNA–mRNA interaction pairs predicted in this study most likely participate in the regulation of muscle development. The genes in these two pairs were *ACTN4*, a transcriptional regulator of myocyte enhancer factor that is associated with skeletal muscle differentiation (An et al., 2014), and *Prox1*, an essential gene for satellite cell differentiation and muscle fiber-type regulation (Kivela et al., 2016). They were paired, respectively, with important muscle-associated miRNAs ssc-miR-133a-3p (Wang et al., 2018c) and ssc-miR-338 (Mcdaneld et al., 2009). Although most of these miRNA–mRNA pairs have been reported to participate in muscle development, the actual interactions between them were still unclear and thus in need of further validation. The data from our study provide a rich resource for determining key interactions of miRNA–mRNA in muscle development.

DATA AVAILABILITY

Sequencing data for miRNA have been deposited to Sequence Read Archive at the National Center for Biotechnical Information under accession number PRJNA552780. The RNA-seq data for mRNA were deposited in the Gene Expression Omnibus under accession number GSE73763.

ETHICS STATEMENT

All animal handlings were approved by the Ethical Committee of the Faculty of Veterinary Medicine (EC2013/118) of Ghent University. All methods were performed in accordance with the relevant guidelines and regulations.

AUTHOR CONTRIBUTIONS

ZT and KL designed the experiment and wrote and revised the manuscript. MC and YaY wrote the manuscript and analyzed the data. YiY completed dual luciferase reporter assay. YT and SL were involved in sample collection and total RNA extraction. MZ installed the software used in this study.

FUNDING

This work was supported by the National Natural Science Foundation of China (31830090), the National Key Project (2016ZX08009-003-006), the Shenzhen Science, Technology and Innovation Commission (JCYJ20170307160516413), the Special Fund for Industrial Development of Dapeng New Area at Shenzhen (KY20180114), and the Agricultural Science and Technology Innovation Program (ASTIP-AGIS5). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00756/full#supplementary-material

REFERENCES

An, H. T., Kim, J., Yoo, S., and Ko, J. (2014). Small leucine zipper protein (sLZIP) negatively regulates skeletal muscle differentiation *via* interaction with alpha-actinin-4. *J. Biol. Chem.* 289, 4969–4979. doi: 10.1074/jbc.M113.515395

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Arora, A., and Dey, C. S. (2014). SIRT2 negatively regulates insulin resistance in C2C12 skeletal muscle cells. *Biochim. Biophys. Acta* 1842, 1372–1378. doi: 10.1016/j.bbadis.2014.04.027

Bachmann, C., Jungbluth, H., Muntoni, F., Manzur, A. Y., Zorzato, F., and Treves, S. (2017). Cellular, biochemical and molecular changes in muscles from patients with X-linked myotubular myopathy due to MTM1 mutations. *Hum. Mol. Genet.* 26, 320–332. doi: 10.1093/hmg/ddw388

Bai, L., Liang, R., Yang, Y., Hou, X., Wang, Z., Zhu, S., et al. (2015). MicroRNA-21 regulates PI3K/Akt/mTOR signaling by targeting TGFbetaI during skeletal

muscle development in pigs. *PLoS One* 10, e0119396. doi: 10.1371/journal.pone.0119396

Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B. D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Genes Dev.* 26, 2763–2779. doi: 10.1101/gad.200113.112

Burguiere, A. C., Nord, H., and Von Hofsten, J. (2011). Alkali-like myosin light chain-1 (myl1) is an early marker for differentiating fast muscle cells in zebrafish. *Dev. Dyn.* 240, 1856–1863. doi: 10.1002/dvdy.22677

Chakroun, I., Yang, D., Girgis, J., Gunasekharan, A., Phenix, H., Kaern, M., et al. (2015). Genome-wide association between Six4, MyoD, and the histone demethylase Utx during myogenesis. *FASEB J.* 29, 4738–4755. doi: 10.1096/fj.15-277053

Chao, Z., Wu, J., Zheng, R., Li, F. E., Xiong, Y. Z., and Deng, C. Y. (2011). Molecular characterization and expression patterns of Lbx1 in porcine skeletal muscle. *Mol. Biol. Rep.* 38, 3983–3991. doi: 10.1007/s11033-010-0516-1

Chen, G., Masuda, A., Konishi, H., Ohkawara, B., Ito, M., Kinoshita, M., et al. (2016). Phenylbutazone induces expression of MBNL1 and suppresses formation of MBNL1-CUG RNA foci in a mouse model of myotonic dystrophy. *Sci. Rep.* 6, 25317. doi: 10.1038/srep25317

Collard, L., Herledan, G., Pincini, A., Guerci, A., Randrianarison-Huetz, V., and Sotiropoulos, A. (2014). Nuclear actin and myocardin-related transcription factors control disuse muscle atrophy through regulation of Srf activity. *J. Cell. Sci.* 127, 5157–5163. doi: 10.1242/jcs.155911

Dabaja, A. A., Mielnik, A., Robinson, B. D., Wosnitzer, M. S., Schlegel, P. N., and Paduch, D. A. (2015). Possible germ cell-Sertoli cell interactions are critical for establishing appropriate expression levels for the Sertoli cell-specific microRNA, miR-202-5p, in human testis. *Basic Clin. Androl.* 25, 2. doi: 10.1186/s12610-015-0018-z

Ding, X., Zhou, S., Li, M., Cao, C., Wu, P., Sun, L., et al. (2017). Upregulation of SRF is associated with hypoxic pulmonary hypertension by promoting viability of smooth muscle cells *via* increasing expression of Bcl-2. *J. Cell. Biochem.* 118, 2731–2738. doi: 10.1002/jcb.25922

Dunaeva, M., Blom, J., Thurlings, R., and Pruijn, G. J. M. (2018). Circulating serum miR-223-3p and miR-16-5p as possible biomarkers of early rheumatoid arthritis. *Clin. Exp. Immunol.* 193, 376–385. doi: 10.1111/cei.13156

Emrani, R., Rebillard, A., Lefeuvre, L., Gratas-Delamarche, A., Davies, K. J., and Cillard, J. (2015). The calcineurin antagonist RCAN1-4 is induced by exhaustive exercise in rat skeletal muscle. *Free Radic. Biol. Med.* 87, 290–299. doi: 10.1016/j.freeradbiomed.2015.06.023

Fortes, M. A., Marzuca-Nassr, G. N., Vitzel, K. F., Da Justa Pinheiro, C. H., Newsholme, P., and Curi, R. (2016). Housekeeping proteins: how useful are they in skeletal muscle diabetes studies and muscle hypertrophy models? *Anal. Biochem.* 504, 38–40. doi: 10.1016/j.ab.2016.03.023

Fougerousse, F., Durand, M., Suel, L., Pourquie, O., Delezoide, A. L., Romero, N. B., et al. (1998). Expression of genes (CAPN3, SGCA, SGCB, and TTN) involved in progressive muscular dystrophies during early human development. *Genomics* 48, 145–156. doi: 10.1006/geno.1997.5160

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Giuffra, E., Kijas, J. M., Amarger, V., Carlborg, O., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785–1791.

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622

Hashemi, M. S., Mozdarani, H., Ghaedi, K., and Nasr-Esfahani, M. H. (2018). Among seven testis-specific molecular markers, SPEM1 appears to have a significant clinical value for prediction of sperm retrieval in azoospermic men. *Andrology* 6, 890–895. doi: 10.1111/andr.12528

Herrmann, J., Borkham-Kamphorst, E., Haas, U., Van De Leur, E., Fraga, M. F., Esteller, M., et al. (2006). The expression of CSRP2 encoding the LIM domain protein CRP2 is mediated by TGF-beta in smooth muscle and hepatic stellate cells. *Biochem. Biophys. Res. Commun.* 345, 1526–1535. doi: 10.1016/j.bbrc.2006.05.076

Hou, X., Tang, Z., Liu, H., Wang, N., Ju, H., and Li, K. (2012). Discovery of microRNAs associated with myogenesis by deep sequencing of serial developmental skeletal muscles in pigs. *PLoS One* 7, e52123. doi: 10.1371/journal.pone.0052123

Hou, X., Yang, Y., Zhu, S., Hua, C., Zhou, R., Mu, Y., et al. (2016). Comparison of skeletal muscle miRNA and mRNA profiles among three pig breeds. *Mol. Genet. Genomics* 291, 559–573. doi: 10.1007/s00438-015-1126-3

Hu, Q., Tong, H., Zhao, D., Cao, Y., Zhang, W., Chang, S., et al. (2015). Generation of an efficient artificial promoter of bovine skeletal muscle alpha-actin gene (ACTA1) through addition of cis-acting element. *Cell. Mol. Biol. Lett.* 20, 160–176. doi: 10.1515/cmble-2015-0009

Huang Da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., et al. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35, W169–W175. doi: 10.1093/nar/gkm415

Iwakiri, J., Terai, G., and Hamada, M. (2017). Computational prediction of lncRNA–mRNA interactionsby integrating tissue specificity in human transcriptome. *Biol. Direct* 12, 15. doi: 10.1186/s13062-017-0183-4

Kaminski, M. J., Kaminska, M., Skorupa, I., Kazimierczyk, R., Musial, W. J., and Kaminski, K. A. (2013). In-silico identification of cardiovascular disease-related SNPs affecting predicted microRNA target sites. *Pol. Arch. Med. Wewn.* 123, 355–363. doi: 10.20452/pamw.1819

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36

Kim, P., Park, A., Han, G., Sun, H., Jia, P., and Zhao, Z. (2018). TissGDB: tissue-specific gene database in cancer. *Nucleic Acids Res.* 46, D1031–D1038. doi: 10.1093/nar/gkx850

Kivela, R., Salmela, I., Nguyen, Y. H., Petrova, T. V., Koistinen, H. A., Wiener, Z., et al. (2016). The transcription factor Prox1 is essential for satellite cell differentiation and muscle fibre-type regulation. *Nat. Commun.* 7, 13124. doi: 10.1038/ncomms13124

Koh, W., Pan, W., Gawad, C., Fan, H. C., Kerchner, G. A., Wyss-Coray, T., et al. (2014). Noninvasive *in vivo* monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7361–7366. doi: 10.1073/pnas.1405528111

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Kruger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34, W451–W454. doi: 10.1093/nar/gkl243

Lage, K., Hansen, N. T., Karlberg, E. O., Eklund, A. C., Roque, F. S., Donahoe, P. K., et al. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20870–20875. doi: 10.1073/pnas.0810772105

Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., et al. (2017). A comprehensive mouse transcriptomic bodymap across 17 tissues by RNA-seq. *Sci. Rep.* 7, 4200. doi: 10.1038/s41598-017-04520-z

Li, J. J., Huang, H., Bickel, P. J., and Brenner, S. E. (2014). Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res.* 24, 1086–1101. doi: 10.1101/gr.170100.113

Liang, G., Yang, Y., Niu, G., Tang, Z., and Li, K. (2017). Genome-wide profiling of *Sus scrofa* circular RNAs across nine organs and three developmental stages. *DNA Res.* 24, 523–535. doi: 10.1093/dnares/dsx022

Lin, H. Y., Cheng, C. H., Chen, D. T., Chen, Y. A., and Park, J. Y. (2016). Coexpression and expression quantitative trait loci analyses of the angiogenesis gene–gene interaction network in prostate cancer. *Transl. Cancer Res.* 5, S951–S963. doi: 10.21037/tcr.2016.10.55

Liu, H., Xi, Y., Liu, G., Zhao, Y., Li, J., and Lei, M. (2018). Comparative transcriptomic analysis of skeletal muscle tissue during prenatal stages in Tongcheng and Yorkshire pig using RNA-seq. *Funct. Integr. Genomics* 18, 195–209. doi: 10.1007/s10142-017-0584-6

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Mcdaneld, T. G., Smith, T. P., Doumit, M. E., Miles, J. R., Coutinho, L. L., Sonstegard, T. S., et al. (2009). MicroRNA transcriptome profiles during swine skeletal muscle development. *BMC Genomics* 10, 77. doi: 10.1186/1471-2164-10-77

Mcloughlin, K. E., Nalpas, N. C., Rue-Albrecht, K., Browne, J. A., Magee, D. A., Killick, K. E., et al. (2014). RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis*. *Front. Immunol.* 5, 396. doi: 10.3389/fimmu.2014.00396

Mora, A., Davies, A. M., Bertrand, L., Sharif, I., Budas, G. R., Jovanovic, S., et al. (2003). Deficiency of PDK1 in cardiac muscle results in heart failure and increased sensitivity to hypoxia. *EMBO J.* 22, 4666–4676. doi: 10.1093/emboj/cdg469

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., et al. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43, D130–D137. doi: 10.1093/nar/gku1063

Neville, M. J., Collins, J. M., Gloyn, A. L., Mccarthy, M. I., and Karpe, F. (2011). Comprehensive human adipose tissue mRNA and microRNA endogenous control selection for quantitative real-time-PCR normalization. *Obesity (Silver Spring)* 19, 888–892. doi: 10.1038/oby.2010.257

Nielsen, M., Hansen, J. H., Hedegaard, J., Nielsen, R. O., Panitz, F., Bendixen, C., et al. (2010). MicroRNA identity and abundance in porcine skeletal muscles determined by deep sequencing. *Anim. Genet.* 41, 159–168. doi: 10.1111/j.1365-2052.2009.01981.x

Park, M., Lee, B. S., Jeon, S. H., Nam, H. J., Lee, G., Kim, C. H., et al. (2015). A novel isoform of met receptor tyrosine kinase blocks hepatocyte growth factor/Met signaling and stimulates skeletal muscle cell differentiation. *J. Biol. Chem.* 290, 1804–1817. doi: 10.1074/jbc.M114.596957

Peng, Y., Xiang, H., Chen, C., Zheng, R., Chai, J., Peng, J., et al. (2013). MiR-224 impairs adipocyte early differentiation and regulates fatty acid metabolism. *Int. J. Biochem. Cell. Biol.* 45, 1585–1593. doi: 10.1016/j.biocel.2013.04.029

Ran, M., Chen, B., Wu, M., Liu, X., He, C., Yang, A., et al. (2015). Integrated analysis of miRNA and mRNA expression profiles in development of porcine testes. *RSC Adv.* 5, 63439–63449. doi: 10.1039/C5RA07488F

Ran, M., Li, Z., Cao, R., Weng, B., Peng, F., He, C., et al. (2018a). miR-26a suppresses autophagy in swine Sertoli cells by targeting ULK2. *Reprod. Domest. Anim.* 53, 864–871. doi: 10.1111/rda.13177

Ran, M., Weng, B., Cao, R., Li, Z., Peng, F., Luo, H., et al. (2018b). miR-26a inhibits proliferation and promotes apoptosis in porcine immature Sertoli cells by targeting the PAK2 gene. *Reprod. Domest. Anim.* 53, 1375–1385. doi: 10.1111/rda.13254

Riffo-Campos, A. L., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for sequence-based miRNA target prediction: what to choose? *Int. J. Mol. Sci.* 17, E1987. doi: 10.3390/ijms17121987

Roux, J., Gonzalez-Porta, M., and Robinson-Rechavi, M. (2012). Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucleic Acids Res.* 40, 5890–5900. doi: 10.1093/nar/gks279

Runfola, V., Sebastian, S., Dilworth, F. J., and Gabellini, D. (2015). Rbfox proteins regulate tissue-specific alternative splicing of Mef2D required for muscle differentiation. *J. Cell. Sci.* 128, 631–637. doi: 10.1242/jcs.161059

Sakellariou, G. K., Mcdonagh, B., Porter, H., Giakoumaki, I. I., Earl, K. E., et al. (2018). Comparison of whole body SOD1 knockout with muscle-specific SOD1 knockout mice reveals a role for nerve redox signaling in regulation of degenerative pathways in skeletal muscle. *Antioxid. Redox Signal.* 28, 275–295. doi: 10.1089/ars.2017.7249

Santos, C. A., Blanck, D. V., and De Freitas, P. D. (2014). RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Front. Genet.* 5, 298. doi: 10.3389/fgene.2014.00298

Sato, T., Yamamoto, T., and Sehara-Fujisawa, A. (2014). miR-195/497 induce postnatal quiescence of skeletal muscle stem cells. *Nat. Commun.* 5, 4597. doi: 10.1038/ncomms5597

Shen, S., Li, D., Liang, J., and Wang, J. (2019). Testis-specific calcium-binding protein CBP86-IV (CABYR) binds with phosphoglycerate kinase 2 *in vitro* and *in vivo* experiment. *Andrologia* 10, e13287. doi: 10.1111/and.13287

Siengdee, P., Trakooljul, N., Murani, E., Schwerin, M., Wimmers, K., and Ponsuksili, S. (2015). MicroRNAs regulate cellular ATP levels by targeting mitochondrial energy metabolism genes during C2C12 myoblast differentiation. *PLoS One* 10, e0127850. doi: 10.1371/journal.pone.0127850

Sromek, M., Glogowski, M., Chechlinska, M., Kulinczak, M., Szafron, L., Zakrzewska, K., et al. (2017). Changes in plasma miR-9, miR-16, miR-205 and miR-486 levels after non-small cell lung cancer resection. *Cell. Oncol. (Dordr)* 40, 529–536. doi: 10.1007/s13402-017-0334-8

Staus, D. P., Blaker, A. L., Medlin, M. D., Taylor, J. M., and Mack, C. P. (2011). Formin homology domain-containing protein 1 regulates smooth muscle cell phenotype. *Arterioscler. Thromb. Vasc. Biol.* 31, 360–367. doi: 10.1161/ATVBAHA.110.212993

Szabo, L., Morey, R., Palpant, N. J., Wang, P. L., Afari, N., Jiang, C., et al. (2015). Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol.* 16, 126. doi: 10.1186/s13059-015-0690-5

Tallquist, M. D., Weismann, K. E., Hellstrom, M., and Soriano, P. (2000). Early myotome specification regulates PDGFA expression and axial skeleton development. *Development* 127, 5059–5070.

Tang, Z., Wu, Y., Yang, Y., Yang, Y. T., Wang, Z., Yuan, J., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutional conservation in Sus scrofa. *Sci. Rep.* 7, 43166. doi: 10.1038/srep43166

Tang, Z., Yang, Y., Wang, Z., Zhao, S., Mu, Y., and Li, K. (2015). Integrated analysis of miRNA and mRNA paired expression profiling of prenatal skeletal muscle development in three genotype pigs. *Sci. Rep.* 5, 15544. doi: 10.1038/srep15544

Villmow, M., Klockner, U., Heymes, C., Gekle, M., and Rueckschloss, U. (2015). NOS1 induces NADPH oxidases and impairs contraction kinetics in aged murine ventricular myocytes. *Basic Res. Cardiol.* 110, 506. doi: 10.1007/s00395-015-0506-5

Wang, H., Zhang, P., Chen, W., Feng, D., Jia, Y., and Xie, L. X. (2012a). Evidence for serum miR-15a and miR-16 levels as biomarkers that distinguish sepsis from systemic inflammatory response syndrome in human subjects. *Clin. Chem. Lab. Med.* 50, 1423–1428. doi: 10.1515/cclm-2011-0826

Wang, X., Du, C., He, X., Deng, X., He, Y., and Zhou, X. (2018a). MiR-4463 inhibits the migration of human aortic smooth muscle cells by AMOT. *Biosci. Rep.* 38, BSR20180150. doi: 10.1042/BSR20180150

Wang, X., Zhang, P., Li, L., Che, D., Li, T., Li, H., et al. (2018b). miRNA editing landscape reveals miR-34c regulated spermatogenesis through structure and target change in pig and mouse. *Biochem. Biophys. Res. Commun.* 502, 486–492. doi: 10.1016/j.bbrc.2018.05.197

Wang, X. G., Yu, J. F., Zhang, Y., Gong, D. Q., and Gu, Z. L. (2012b). Identification and characterization of microRNA from chicken adipose tissue and skeletal muscle. *Poult. Sci.* 91, 139–149. doi: 10.3382/ps.2011-01656

Wang, Y., Ma, J., Qiu, W., Zhang, J., Feng, S., Zhou, X., et al. (2018c). Guanidinoacetic acid regulates myogenic differentiation and muscle growth through miR-133a-3p and miR-1a-3p co-mediated Akt/mTOR/S6K signaling pathway. *Int. J. Mol. Sci.* 19, E2837. doi: 10.3390/ijms19092837

Wu, Y., Li, Z., Yang, M., Dai, B., Hu, F., Yang, F., et al. (2017). MicroRNA-214 regulates smooth muscle cell differentiation from stem cells by targeting RNA-binding protein QKI. *Oncotarget* 8, 19866–19878. doi: 10.18632/oncotarget.15189

Yang, J. C., Wu, S. C., Rau, C. S., Chen, Y. C., Lu, T. H., Wu, Y. C., et al. (2015). TLR4/NF-kappaB-responsive microRNAs and their potential target genes: a mouse model of skeletal muscle ischemia–reperfusion injury. *Biomed. Res. Int.* 2015, 410721. doi: 10.1155/2015/410721

Yang, Y., Liang, G., Niu, G., Zhang, Y., Zhou, R., Wang, Y., et al. (2017). Comparative analysis of DNA methylome and transcriptome of skeletal muscle in lean-, obese-, and mini-type pigs. *Sci. Rep.* 7, 39883. doi: 10.1038/srep39883

Yang, Y., Zhou, R., Mu, Y., Hou, X., Tang, Z., and Li, K. (2016). Genome-wide analysis of DNA methylation in obese, lean, and miniature pig breeds. *Sci. Rep.* 6, 30160. doi: 10.1038/srep30160

Yu, Y., Fuscoe, J. C., Zhao, C., Guo, C., Jia, M., Qing, T., et al. (2014). A rat RNA-seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat. Commun.* 5, 3230. doi: 10.1038/ncomms4230

Yuan, H., Niu, Y., Liu, X., and Fu, L. (2014). Exercise increases the binding of MEF2A to the Cpt1b promoter in mouse skeletal muscle. *Acta Physiol. (Oxf)* 212, 283–292. doi: 10.1111/apha.12395

Zeng, J., Liu, S., Zhao, Y., Tan, X., Aljohi, H. A., Liu, W., et al. (2016). Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. *Gene* 576, 560–570. doi: 10.1016/j.gene.2015.11.003

Zhang, W. R., Zhang, H. N., Wang, Y. M., Dai, Y., Liu, X. F., Li, X., et al. (2017). miR-143 regulates proliferation and differentiation of bovine skeletal muscle satellite cells by targeting IGFBP5. *In Vitro Cell. Dev. Biol. Anim.* 53, 265–271. doi: 10.1007/s11626-016-0109-y

Zhang, X., Azhar, G., Williams, E. D., Rogers, S. C., and Wei, J. Y. (2015a). MicroRNA clusters in the adult mouse heart: age-associated changes. *Biomed. Res. Int.* 2015, 732397. doi: 10.1155/2015/732397

Zhang, Y., Jing, J., Li, X., Wang, J., Feng, X., Cao, R., et al. (2015b). Integration analysis of miRNA and mRNA expression profiles in swine testis cells infected with Japanese encephalitis virus. *Infect. Genet. Evol.* 32, 342–347. doi: 10.1016/j.meegid.2015.03.037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Yao, Yang, Zhu, Tang, Liu, Li and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 September 2019
doi: 10.3389/fgene.2019.00863

Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep

Mazdak Salavati[1], Stephen J. Bush[1], Sergio Palma-Vera[2], Mary E. B. McCulloch[1], David A. Hume[3] and Emily L. Clark[1]**

[1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh, United Kingdom, [2] Leibniz Institute for Farm Animal Biology (FBN), Institute for Reproductive Biology, Dummerstorf, Germany, [3] Mater Research Institute-University of Queensland, Translational Research Institute, Woolloongabba, QLD, Australia

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Amanda Chamberlain, Agriculture Victoria, Australia
Dan Nonneman, United States Department of Agriculture, United States

***Correspondence:**
Mazdak Salavati
Mazdak.Salavati@roslin.ed.ac.uk
Emily L. Clark
Emily.Clark@roslin.ed.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *25 April 2019*
Accepted: *19 August 2019*
Published: *19 September 2019*

Citation:
Salavati M, Bush SJ, Palma-Vera S, McCulloch MEB, Hume DA and Clark EL (2019) Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep. Front. Genet. 10:863. doi: 10.3389/fgene.2019.00863

Pervasive allelic variation at both gene and single nucleotide level (SNV) between individuals is commonly associated with complex traits in humans and animals. Allele-specific expression (ASE) analysis, using RNA-Seq, can provide a detailed annotation of allelic imbalance and infer the existence of cis-acting transcriptional regulation. However, variant detection in RNA-Seq data is compromised by biased mapping of reads to the reference DNA sequence. In this manuscript, we describe an unbiased standardized computational pipeline for allele-specific expression analysis using RNA-Seq data, which we have adapted and developed using tools available under open license. The analysis pipeline we present is designed to minimize reference bias while providing accurate profiling of allele-specific expression across tissues and cell types. Using this methodology, we were able to profile pervasive allelic imbalance across tissues and cell types, at both the gene and SNV level, in Texel×Scottish Blackface sheep, using the sheep gene expression atlas data set. ASE profiles were pervasive in each sheep and across all tissue types investigated. However, ASE profiles shared across tissues were limited, and instead, they tended to be highly tissue-specific. These tissue-specific ASE profiles may underlie the expression of economically important traits and could be utilized as weighted SNVs, for example, to improve the accuracy of genomic selection in breeding programs for sheep. An additional benefit of the pipeline is that it does not require parental genotypes and can therefore be applied to other RNA-Seq data sets for livestock, including those available on the Functional Annotation of Animal Genomes (FAANG) data portal. This study is the first global characterization of moderate to extreme ASE in tissues and cell types from sheep. We have applied a robust methodology for ASE profiling to provide both a novel analysis of the multi-dimensional sheep gene expression atlas data set and a foundation for identifying the regulatory and expressed elements of the genome that are driving complex traits in livestock.

Keywords: allele-specific expression, mapping bias, RNA-Seq, sheep, transcriptome, WASP, GeneiASE

INTRODUCTION

Allele-specific expression (ASE) is the imbalance of allelic expression between parental (diploid) copies at the same locus (Barlow and Bartolomei, 2014). It is most commonly associated with *cis*-acting regulatory variation that may mediate parent-of-origin, sex- or tissue-specific transcription of one allele relative to the other (Renfree et al., 2009; Hasin-Brumshtein et al., 2014). In a single individual, where there are informative sequence variants (i.e., heterozygote loci) that distinguish the products of two alleles, ASE can be detected by RNA sequencing (Chamberlain et al., 2015; GTEx Consortium et al., 2017; Cao et al., 2019; Guillocheau et al., 2019). The ratio of allelic read counts obtained from RNA-Seq data sets can be used as a reliable proxy for ASE [i.e., $ASEratio = CountsAllele1/(CountsAllele1 + CountsAllele2)$] (Edsgärd et al., 2016).

Large and complex RNA-Seq data sets give rise to unique and interesting computational challenges, in particular the elimination of reference mapping bias in ASE analysis of diploid genomes. RNA-Seq data are commonly mapped against reference genomes which are typically "flat," with each position represented only by the reference (most abundant) allele. As such, reads containing heterozygous loci are more likely to be erroneously mapped (Degner et al., 2009; Stevenson et al., 2013; Hodgkinson et al., 2016). This can lead to high false-positive ASE locus discovery rates (Degner et al., 2009). Although development of *de novo* transcript assemblers (Zerbino and Birney, 2008), usage of personalized reference genomes (Rozowsky et al., 2011; Smith et al., 2013), variant-aware aligners (Xin et al., 2013; Hach et al., 2014), and mapping-free quantification e.g., Kallisto (Bray et al., 2016) have resolved some of these issues, reference allele mapping bias remains a considerable challenge in ASE studies. In the absence of "trios" of animals or reference population phased haplotype information, which are rare for livestock, correction of mapping bias *via* synthetic reads with either N masking or alternative mapping bias correction at the heterozygote sites, has proven a robust alternative for ASE discovery (Degner et al., 2009; Mayba et al., 2014; van de Geijn et al., 2015; Miao et al., 2018). In 2015, Van de Geijn et al. benchmarked the WASP software mapping correction strategy against N-masked reads and personal genome mapping. WASP showed consistent correct mapping of reads with multiple alleles and lower false discovery rates (FDR) in comparison to the other two methods (van de Geijn et al., 2015). The analysis pipeline we present in this manuscript is based on WASP's methodology and is designed to minimize reference bias while providing accurate profiling of allele-specific expression in large and complex RNA-Seq data sets.

We have developed an ASE analysis pipeline using the combination of software available under open license, WASP (reference mapping bias removal) (van de Geijn et al., 2015), GATK (ASEReadCounter) (McKenna et al., 2010; Van der Auwera et al., 2013), and GeneiASE (Liptak-Stouffer aggregative ASE gene model) (Edsgärd et al., 2016). The GeneiASE model is capable of testing ASE at the gene level using two approaches: i) static ASE, which measures allelic imbalance within a gene (i.e., when ASE variants are located within the boundaries of the gene); and ii) individual condition-dependent ASE (ICD), which measures inducible ASE in a gene under an environmental pressure between two timepoints (i.e., in stimulated or unstimulated immune cells).

In addition to ASE at the gene level, we can also measure significant ASE at the single-nucleotide level (SNV). ASE has been shown to be enriched within expression quantitative trait loci (eQTL) regions (Montgomery et al., 2010); therefore, identifying ASE variants can be useful for understanding the transcriptomic control of complex traits in livestock. Complex trait mapping of ASE loci has been associated with phenotypes, such as resistance to Marek's disease in chicken (Meydan et al., 2011) and pigmentation patterns in sheep (García-Gámez et al., 2011).

Understanding ASE is also important because cross-breeding now underlies most livestock production systems. Knowledge of ASE may provide insights into the molecular basis of the complex phenomenon of hybrid vigor, as emphasized by recent studies on two Chinese goat breeds and their F1 hybrids (Cao et al., 2019) and in F1 crosses of two highly inbred chicken lines (Zhuo et al., 2017). In this study, we measure ASE in crossbred sheep. Sheep are an economically important livestock species in many countries across the globe and particularly in emerging economies. The identification of prevalent ASE in populations or breeds, especially in economically relevant phenotypes and tissues could be used to improve genomic prediction in sheep breeding programs, such as those that have been established in Australia and New Zealand (Daetwyler et al., 2010).

Using the methodology we describe, for mapping bias correction and robust positive ASE discovery, we were able to profile pervasive allelic imbalance across tissues and cell types, at both the gene and SNV level, in Texel×Scottish Blackface sheep. We analyzed a subset of total RNA-Seq libraries from liver, spleen, ileum, thymus, and bone marrow-derived macrophages (BMDM) (±) lipopolysaccharide (LPS) from six individual adult crossbred sheep to produce a detailed picture of allelic imbalance in immune-related tissues and cell types. We chose to focus this analysis on immune-related tissues in part because of the depth of available sequence in those tissues, and in part because they contain abundant immune cell populations. The diversity of cell populations is reflected in the transcriptional complexity of immune tissues and cell types in the sheep gene expression atlas data set (Clark et al., 2017; Bush et al., 2019). As such, this subset of tissues gave us a transcriptionally rich data set in which to measure ASE. We also included BMDMs stimulated and unstimulated with LPS to mimic infection with Gram-negative bacteria to test whether ASE changed in response to stimulation with LPS in these cells. By measuring ASE in these tissues and cell types from sheep we were able to: i) provide insight into how pervasive ASE is across tissues at the gene and SNV level, ii) generate tissue-specific ASE profiles, iii) investigate sex-specific patterns of ASE, and iv) determine the extent to which ASE changes in response to stimulation with LPS in an immune cell type. This novel analysis of the multi-dimensional sheep gene expression atlas data set provides a foundation for further analysis of the regulatory and expressed elements of the genome that are driving complex traits in sheep.

METHODS

Sample Preparation and RNA Extraction

Data from three male and three female Texel×Scottish Blackface (T×BF) sheep from the sheep gene expression atlas project (Clark et al., 2017) were used in this study. The data set including: one cell type (BMDMs (±) LPS treatment) and four tissues (thymus, spleen, liver and ileum). Tissue collection, storage, and RNA extraction are described in Clark et al. (2017). BMDMs were cultured *in vitro* for 7 days in the presence of macrophage colony-stimulating factor (CSF1 (10^4 U/ml)) and unstimulated (0 h –LPS) and stimulated (7 h +100 ng/ml LPS) samples of BMDMs were obtained as previously described (Clark et al., 2017). A total of two samples (one thymus and one spleen) did not pass the RNA quality control (RNA integrity number (Mueller et al., 2004); RIN^e >7) and were not included in the sheep gene expression atlas. Library preparation was performed by Edinburgh Genomics (Edinburgh Genomics, Edinburgh, UK). All total RNA Illumina TruSeq libraries (125 bp paired end) were sequenced at a depth of > 100 million reads per sample.

Reference Mapping Bias Removal

BAM files from RNA-Seq data were previously produced by mapping fastq files to the Oar v3.1 top level DNA fasta track, using HISAT2 (default mismatch penalty MX = 6 MN = 2) as previously described (Clark et al., 2017). Detailed settings and parameters for all the tools used to generate the BAM files can be found at FAANG (2018). These BAM files were used to locate reads with heterozygote loci using WASP's find_intersecting.py script (van de Geijn et al., 2015). The intersection of reads and heterozygote loci in all samples were based on the Ensembl v92 variant call format (VCF) track (Ensembl v92: ovis_aries_incl_consequences.vcf.gz). Briefly, the Ensembl VCF file was filtered for bi-allelic variants within exonic regions, 5k up or downstream of exonic regions (5′ or 3′ UTRs) and intronic regions of all transcripts within the Oar3.1 sheep assembly (exclusion of indels and intergenic variants). These variants were used in WASP's find_intersecting.py script to extract reads mapped to coordinates containing variants for each gene. As a result, reads aligned to exonic, 5′ or 3′ UTRs and intronic regions were separated into reads intersecting heterozygote loci and reads that did not intersect heterozygote loci. Synthetic copies of reads intersecting heterozygote loci were created with the alternate allele flipped to the remaining options of A, T, C, or G [up to 6 loci/read(2^n) max 64 combinations of synthetic reads] using parameters defined in WASP (van de Geijn et al., 2015). This was followed by remapping of the synthetic reads using HISAT2 (default mismatch penalty MX = 6 MN = 2) (Li and Durbin, 2009; Kim et al., 2015) and eliminating the original reads (and their synthetic copies) which mapped to a different coordinate in any of its synthetic copies (WASP's filter_reads.py) (van de Geijn et al., 2015). After merging the retained reads with that did not intersect heterozygote loci, a final BAM file was produced for ASE read counting step (WASP's remove_dup.py).

Allelic Read Counts and Depth Filtration

Allele-specific read counting was carried out using the ASEReadCounter module of GATK v3.8 with parameters -mmq 50 and -mbq 25 (McKenna et al., 2010). Multiple pre-processing steps were performed prior to GeneiASE input as instructed by Edsgärd et al. (2016), which included preparing per chromosome indices, merging the variant set with corresponding gene coordinates, and bi-allelic expression filtering. Loci with < 10 reads mapped were excluded, as were loci with < 3 reads, or < 1% of the total reads, mapped to both the reference and alternative allele. This form of filtration will eliminate loci exhibiting mono-allelic expression (MAE) as previously described (Degner et al., 2009; Stevenson et al., 2013; Mayba et al., 2014). Producing evidence of MAE using total RNA-Seq data sets produced by Illumina short read sequences without parent of origin genotypes or imprinting information has been a controversial issue (DeVeale et al., 2012). Our data set did not include the trios of animals or personalized genomes that would be necessary to resolve MAE. As such, we decided to exclude MAE altogether for our analysis using stringent bi-allelic filtration criteria. Similar bi-allelic filtration criteria have been previously used routinely in ASE studies (Mayba et al., 2014; Chen et al., 2016a; Edsgärd et al., 2016; GTEx Consortium et al., 2017; Raghupathy et al., 2018; Cao et al., 2019; Guillocheau et al., 2019; Gutierrez-Arcelus et al., 2019). The workflow of the analysis pipeline for ASE analysis is detailed in **Figure 1**.

Experimental Design for Defining Allele-Specific Expression

ASE was defined according to the following three categories:

i) Static ASE: which is inherent allelic imbalance (AI) in each gene calculated by ASE at all heterozygote loci (i.e., *ASE = Counts RefAllele*/(*Counts RefAllele + Counts AltAllele*)) within the boundaries of the gene. The effect size of ASE at gene level was produced by aggregation of the ASE effect size at SNVs within the gene boundaries (the Liptak-Stouffer method, applied by the GeneiASE aggregative model). A null distribution of ASE effect size for genes in each transcriptome was produced by random sub-sampling ($n = 1 \times 10^5$) from a pool of genes having min 2 and max 100 loci within their boundaries. The ASE effect size of each gene (aggregated using Liptak-Stouffer) was then tested against the null distribution of the same SNV number, *via* a modified bi-nomial test (2×1 table). Distribution of p values was examined for uniformity prior to FDR correction (**Supplementary Figures S1**, **S2**, and **S3**) (Benjamini and Hochberg, 1995; Pounds and Cheng, 2006; Barton et al., 2013).

ii) Individual condition-dependent ASE (ICD-ASE): in which the same ASE effect size was calculated for each gene in the treated versus the untreated timepoints of the same sample (i.e., BMDM ± LPS). The log2ratio ($ASE_{treated}/ASE_{untreated}$) was used in a beta-binomial test

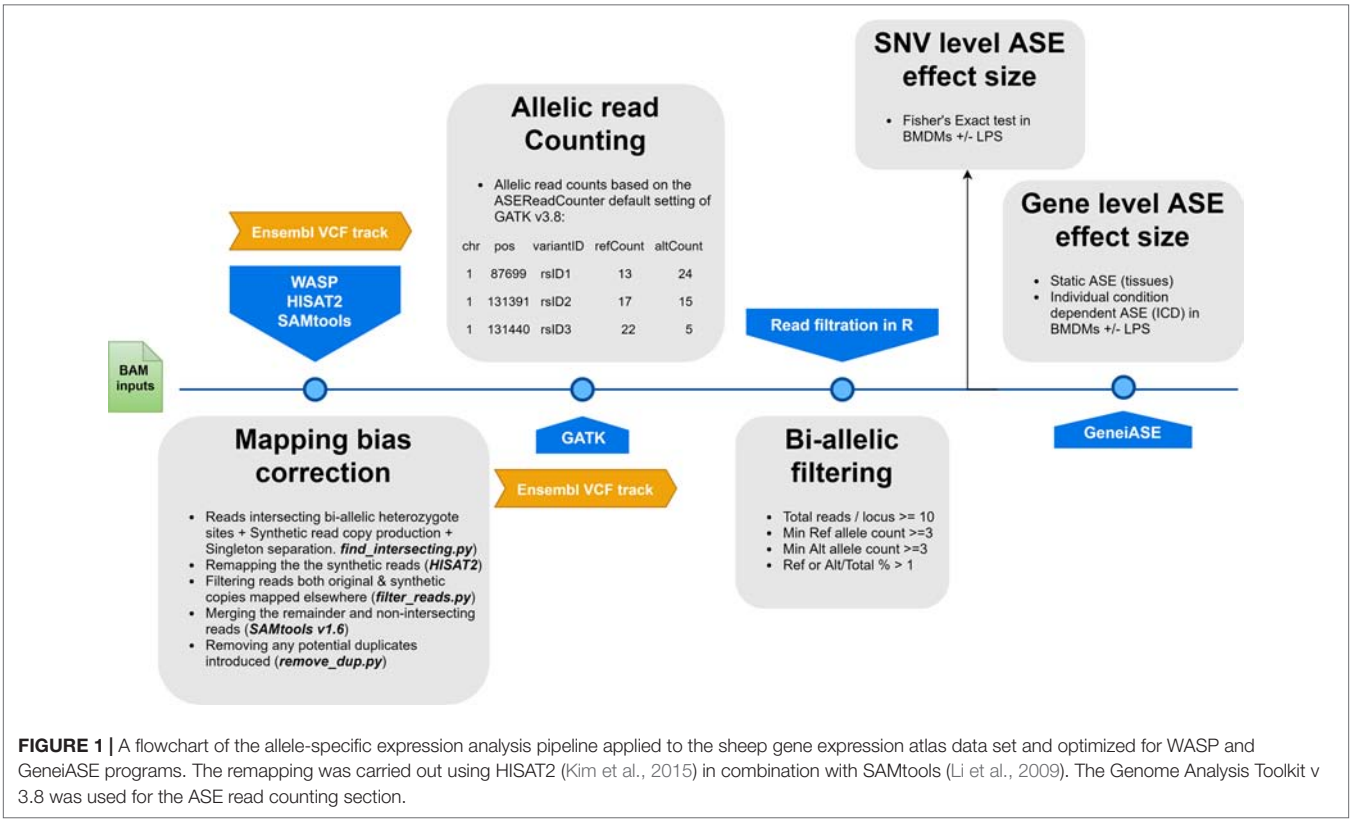


FIGURE 1 | A flowchart of the allele-specific expression analysis pipeline applied to the sheep gene expression atlas data set and optimized for WASP and GeneiASE programs. The remapping was carried out using HISAT2 (Kim et al., 2015) in combination with SAMtools (Li et al., 2009). The Genome Analysis Toolkit v 3.8 was used for the ASE read counting section.

(2×2 table) similar to the static mode. The details of this aggregative model have been previously described in the GeneiASE publication (Edsgärd et al., 2016).

iii) Condition-dependent ASE at SNV level: in which a contingency table was produced for read counts (ref and alt) for every SNV, present both in treated and untreated conditions (BMDM ± LPS) (2×2 table) and a Fisher's exact test performed followed by p value multiple testing correction (Benjamini and Hochberg, 1995). The p values from loci showing ASE and shared by the six adult sheep (ID and coordinate) were unified, using the Stouffer method (Dewey, 2016; Dewey, 2019) and presented as FDR for each locus.

Static ASE was calculated in both tissues and BMDMs (each timepoint was considered separately for BMDMs). Condition-dependent ASE analysis was carried out only in BMDMs ± LPS both at gene (ICD-ASE) and SNV (Fisher's exact) level to study LPS-inducible ASE.

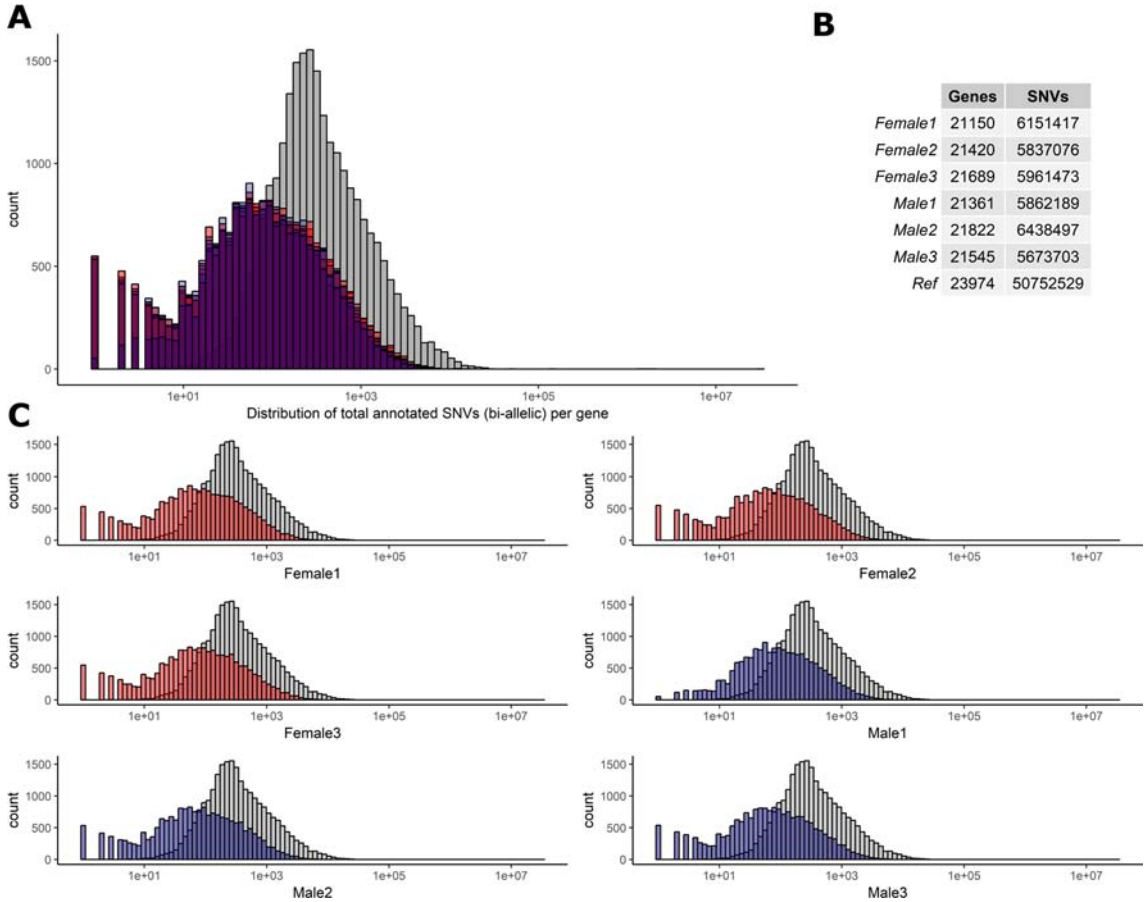
Statistical Analysis and Thresholds Applied

The extraction, transformation and loading of the all data sets and subsequent statistical analysis was carried out in R version 3.4 or higher unless stated otherwise (R Core Team, 2017). System query language join statements (Wickham et al., 2019) were used to compare lists of ASE genes or SNVs between samples. Raw p values resulting from all three types of ASE analysis were corrected for multiple testing *via* Benjamini-Hochberg FDR calculations (Benjamini and Hochberg, 1995). The passing threshold of significance in all analyses was considered to be FDR < 0.1 (10%) except for the Fisher's exact test association study. Genes showing ASE in multiple tissues were considered those for which four or more of the six sheep had significant ASE.

RESULTS

Estimation of Heterozygous Sites Across All Individuals

To determine the level of heterozygosity present in the RNA-Seq data we first assessed the number of bi-allelic heterozygote sites per individual for each of the six sheep (range = 5,673,703–6,438,497) detailed in **Figure 2**. Individual variation was observed in the SNVs per gene in each sheep (**Figures 2A**, **C**). However, there was no significant difference in the total number of bi-allelic SNVs captured in the RNA-Seq data across all six individuals or between the male and female sheep included in the study (**Figure 2B**). The bi-allelic SNVs captured in the RNA-Seq data set were annotated using the Ensembl v.92 (Zerbino et al., 2018) reference VCF track. The distribution of SNVs per gene in the Ensembl track is tail-inflated in comparison to the RNA-Seq data **Figure 2A**. This issue could be due to erroneous assignment of SNVs in hypervariable and repetitive regions, multi-allelic SNVs or



	Genes	SNVs
Female1	21150	6151417
Female2	21420	5837076
Female3	21689	5961473
Male1	21361	5862189
Male2	21822	6438497
Male3	21545	5673703
Ref	23974	50752529

FIGURE 2 | Distribution of biallelic SNVs expressed per gene in each of the six T×BF sheep. The total number of SNVs was averaged across thymus, liver, ileum, and spleen for every animal. Over 5×10^7 SNVs were gathered using Ensembl v.92 VCF track. The total number of SNVs per genes is averaged across four tissue RNA-Seq in each animal (~5.9×10^6). **(A)** Histogram of SNVs per gene counts in the reference track (Ensembl in grey) and six sheep in red (females) and blue (males) overlaid. **(B)**. The overall numbers of genes and SNVs detected in each animal (averaged over four tissues). **(C)** Individual histograms from section A with females in red and males in blue.

simply that there are variants in the Ensembl track that are not expressed (transcribed). The distribution of SNVs for each individual is shown in **Figure 2C**.

Reference Mapping Bias Elimination and Quality Control

We used the WASP ref bias removal script to successfully minimize ref allele mapping bias in the RNA-Seq samples. The mapping bias was assessed by global distribution of the allelic ratio, i.e., $ref_{counts}/alt_{counts} + ref_{counts}$ in each RNA-Seq sample, as shown in **Figure 3** (WASP metrics are included in **Supplementary Figures S10**, **S11**, and **S12**). The ASE discovery rate at the SNV level, on average, constituted 5.8% of the heterozygote loci that passed the minimum filtration criteria in each individual (0.1% of the total expressed). This portion of the transcriptomic variants belonged to an average of 103 genes in each tissue transcriptome (approx. 1%) or 300 in each individual (**Supplementary Figures S5** and **S6**). As shown in **Supplementary Figure S6**, expression level varies across tissues but does not affect the distribution of ASE SNVs.

Genes Exhibiting Tissue-Specific and Pervasive ASE Signatures

We used the static mode of GeneiASE to investigate pervasive and tissue-specific ASE profiles across all of the available samples. Static ASE represents inherent allelic imbalance (AI) in each gene calculated by ASE at all heterozygote loci. The number of genes showing significant static ASE in immune-related tissues across the six sheep are summarized in **Table 1**. On average, approximately 0.5% of the genes in each tissue-specific transcriptome showed significant ASE (approx. 1% of the filtered set of genes). Pervasive ASE genes were investigated by applying the minimum 67% shared rule (i.e., an ASE gene was considered "shared" when it exhibited ASE in a minimum of four of six sheep). A list of ASE genes with significant allelic imbalance (AI) in all tissues, when the effect size was averaged across six sheep, was compiled (**Figure 4A**) (Static ASE measured by GeneiASE's Liptak-Stouffer method). Six genes exhibited pervasive ASE across tissues (i.e., they were shared across all four tissues). In the order of allelic imbalance effect size they were *NAA50*

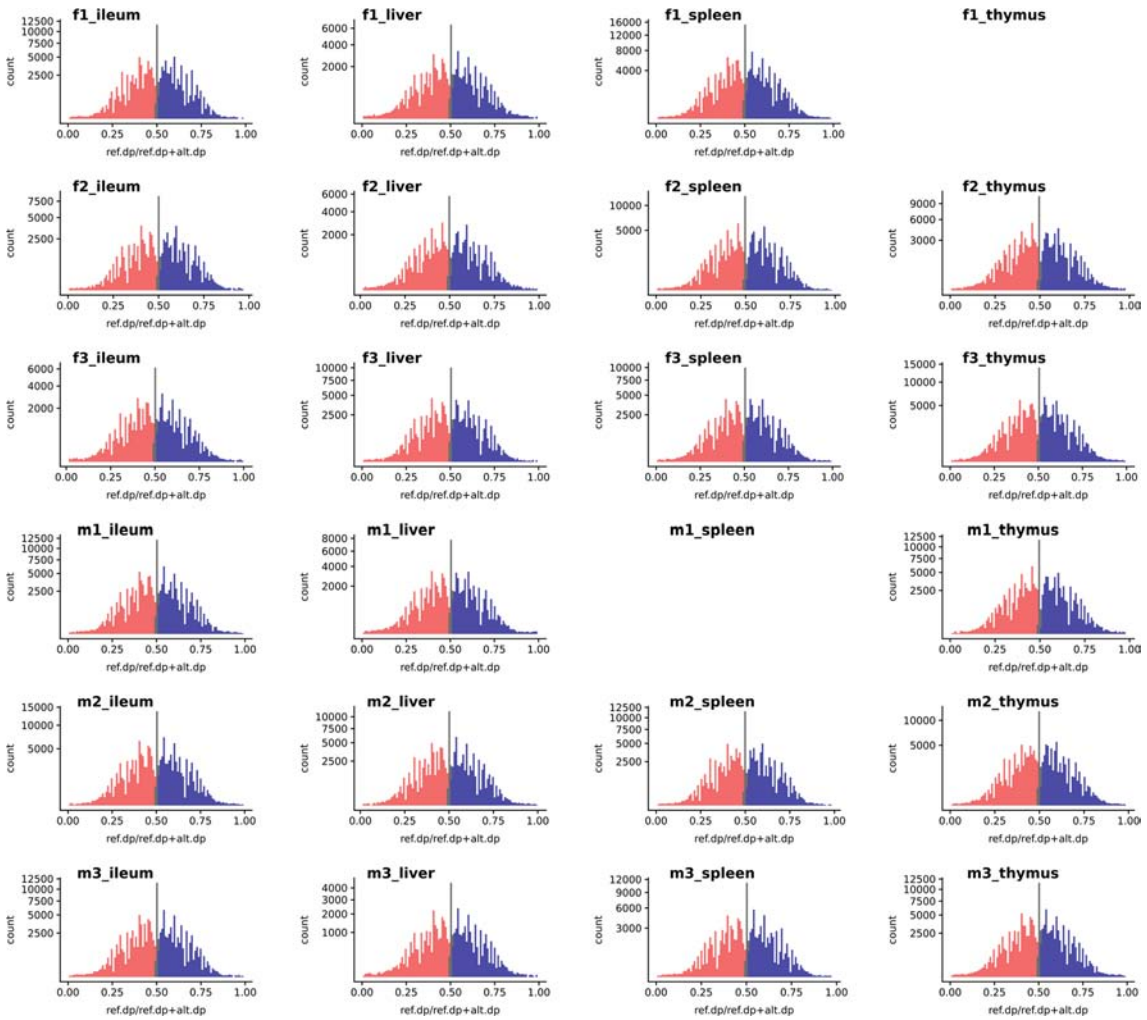


FIGURE 3 | The histogram of a global reference allelic ratio at every locus in the tissues. The distribution of ref allelic ratio showed a balanced profile without any 0 or 1 inflation which is observed in the presence of reference mapping bias. The allelic ratio above 0.51 is shown in blue and below 0.49 in red while balanced bi-allelic expression (0.49—0.51) is colored in gray. Ref.dp, read counts for reference allele; Alt.dp, read counts for alternate allele. The y axis is square root scaled. As discussed in the text SNP that display MAE are not present in any of the samples analyzed, indicating there was no inflation in either 0 or 1 allelic ratio.

TABLE 1 | Total number of genes with significant static ASE in proportion to genes containing informative SNVs (filtered). Total expressed: Average number of genes being expressed in all 4 tissues. Total filtered: Average number of genes (containing heterozygote loci) passing read bi-allelic filtration criteria in 4 tissues. Tissue breakdown has been presented as count (%ASE/filtered).

Sheep	Thymus	Spleen	Liver	Ileum	Total filtered	Total expressed
Female 1	-	136 (1.31%)	153 (1.47%)	92 (0.88%)	10,379	21,150
Female 2	136 (1.28%)	70 (0.66%)	116 (1.09%)	70 (0.66%)	10,572	21,420
Female 3	151 (1.22%)	75 (0.60%)	105 (0.85%)	140 (1.13%)	12,326	21,689
Avg.	143 (1.25%)	94 (0.85%)	125 (1.13%)	101 (0.89%)	11,092	21,419
Male 1	95 (0.92%)	-	157 (1.52%)	86 (0.83%)	10,282	21,361
Male 2	125 (0.99%)	54 (0.43%)	86 (0.68%)	80 (0.63%)	12,514	21,822
Male 3	106 (1.01%)	71 (0.67%)	110 (1.04%)	82 (0.78%)	10,480	21,545
Avg.	109 (0.97%)	62 (0.55%)	118 (1.08%)	83 (0.74%)	11,092	21,576
Total avg.	126 (1.13%)	78 (0.70%)	121 (1.09%)	92 (0.82%)	11,092	21,497

(N(alpha)-acetyltransferase 50, NatE catalytic subunit) with highest ASE effect size in spleen, *UBB* (ubiquitin B) in thymus, *HBP1* (HMG-box transcription factor 1), and *ENSOARG00000016510* both in spleen, *C1orf105* (chromosome 1 open reading frame 105) in ileum and *MTIF2* (mitochondrial translational initiation factor 2) in thymus.

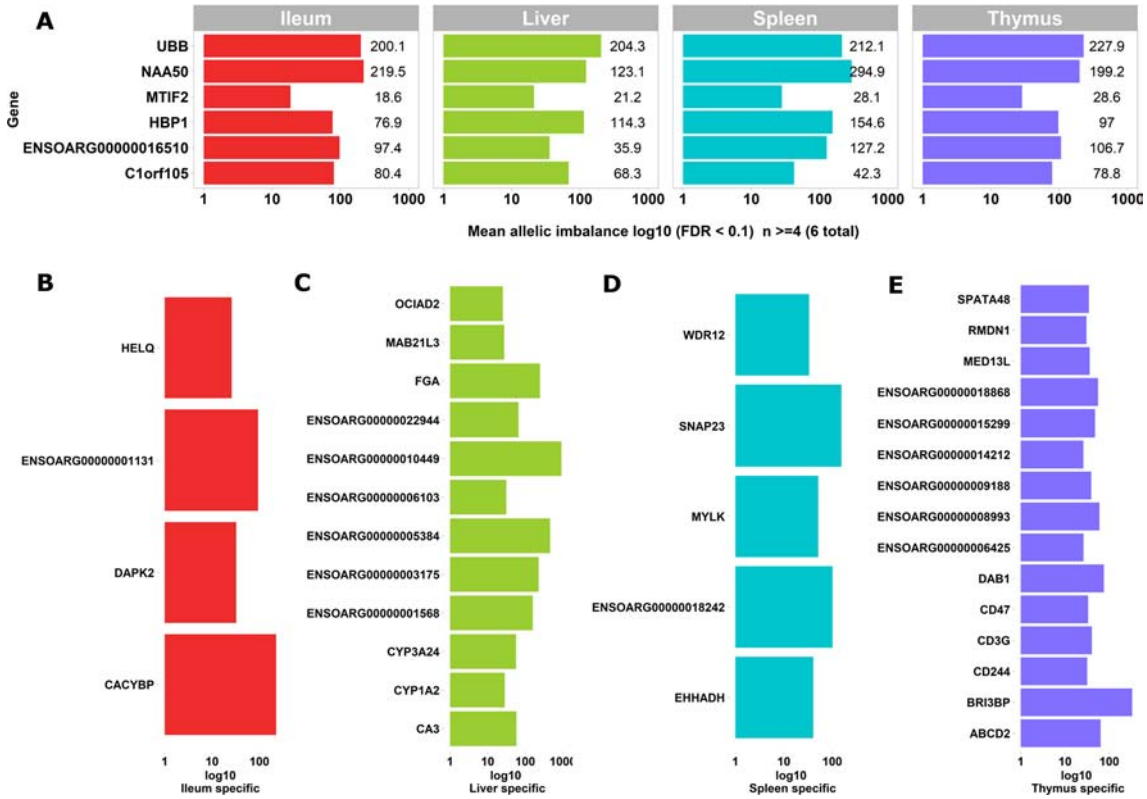


FIGURE 4 | Genes exhibiting static ASE shared across tissues from all six sheep. The x axis represents the mean allelic imbalance (averaged static ASE across sheep in each tissue). **(A)** Genes shared by four tissues with significant (false discovery rate [FDR], < 0.1) static ASE. **(B)** ASE genes private to Ileum. **(C)** Private to liver. **(D)** Private to spleen. **(E)** Private to thymus

Sets of genes with tissue-specific ASE profiles were also captured (**Figures 4B–E**). Thymus had the highest number of tissue-specific ASE genes (n = 15) followed by liver (n = 12), spleen (n = 5), and ileum (n = 4) (**Figures 4B–E**). Among the thymus gene set was *CD244,* which included 30 heterozygote loci with allelic imbalance, one of which was rs406633825. This missense allele (Chr1:110308273 C > A; pVal123Phe MAF = 0.3, SIFT score = 0 deleterious) has previously been reported in the Texel population characterized by the International Sheep Genome Consortium (ISGC) (Kijas et al., 2012). The CD244 protein molecule, a non-MHC (major histocompatibility complex)-mediated marker expressed by NK cells and multiple subsets of CD8+ T cells is known for both pro-inflammatory and inhibitory effects on lymphocytes (McNerney et al., 2005; Georgoudaki et al., 2015). *CD244* exons 2 to 5 are highly conserved in vertebrates and in mouse a trypanosome infection model indicated differential expression was correlated with multiple-copy number variants nearby (Goodhead et al., 2010).The liver-specific ASE profile included genes involved in amino acid metabolism, cytochrome oxidase pathways and fibrinogen: *FGA* (fibrinogen alpha chain), *ENSOARG00000003175* (taurochenodeoxycholic 6 alpha-hydroxylase-like), *ENSOARG00000001568* (novel gene, complement C4-A-like), *CYP3A24* (cytochrome P450 CYP3A24), and *CA3* (carbonic anhydrase 3). Allelic imbalance in spleen was present in *CACYBP* (calcyclin binding protein), *DAPK2* (death-associated protein kinase 2), and a novel gene *GIMAP8*-like (ENSOARG00000001131). The ASE in GIMAP8 has been previously reported in cattle with a strong paternal parent-of-origin expression pattern (Chamberlain et al., 2015). The proteins derived from the *GIMAP/IAN* gene family, are involved in survival, selection, and homeostasis of lymphocytes (Nitta and Takahama, 2007).

Two genes of functional interest showed evidence of strong tissue-specific ASE in the spleen: *SNAP23* and *MYLK*. SNAP23 protein is a key molecule in vesicle transport machinery of the cell and has been reported to be expressed in sheep spleen. SNAP23 or Synaptosome-Associated Protein 23 is part of the protein complex involved in class 1 MHC-mediated antigen processing and presentation and in neutrophil degranulation (Fabregat et al., 2018). *SNAP23* gene is also vital to lymphocyte development (both B and T) *in vitro* (Wong et al., 1997; Kaul et al., 2015). The myosin light chain kinase (*MYLK*) expression in the splenic trabeculae's smooth muscle has been demonstrated previously (Jiang et al., 2014; Clark et al., 2017). Overall, 199 heterozygote bi-allelic loci were present within the *MYLK* gene. The variant rs400678033 (Chr1:186347056G > A;.pAla1014Val), a missense SNV in exon 17 of 33 exons in *MYLK*, showed consistent allelic imbalance in all spleen samples.

In summary, analysis of ASE across immune-related tissues revealed there were a small number of ASE genes that were shared across tissues. ASE signatures instead tended to be tissue-specific, within the sub-set of tissues investigated in this study.

Individual-Specific ASE Signatures

To investigate whether ASE profiles were either shared across all six sheep or private to individual sheep, we used intersectionality (**Figure 5**). Each tissue was investigated separately. A number of private (to each individual) ASE genes were detected for each tissue, ranging from: 31–123 in ileum (**Figure 5A**), 24–80 in liver (**Figure 5B**), 21–83 in spleen (**Figure 5C**), and 31–66 in thymus (**Figure 5D**). Some of the shared sets of ASE genes in these tissues were specific to either male or female sheep, these sex-specific ASE signatures are described in **Figure 5**. In ileum, no sex-specific set was observed (**Figure 5A**). In contrast to ileum, the ASE profile for liver included a single gene with female only membership, *ENSOARG00000017409* (novel gene; 93% orthology with bovine dicarbonyl and L-xylulose reductase [*DCXR*]) (**Figure 5B**). In the spleen, all female sheep shared significant ASE in *PMS1* (PMS1 homolog 1, mismatch repair system component), *ANKRD10* (ankyrin repeat domain 10) and ENSOARG00000006103 which was not present in any of the spleen profiles of male sheep (**Figure 5C**). In the thymus, there were two sex-specific sets: 16 genes showing ASE only in females and five genes only in males. The female-specific thymus gene set included: *ARPP21* (cAMP regulated phosphoprotein 21), *CDKL3* (cyclin-dependent kinase-like 3), *CEP19 (centrosomal protein 19)*, *ENSOARG00000000710* (novel gene), *ENSOARG00000001163* (novel gene), *ENSOARG00000008981* (novel gene; t-lymphocyte surface antigen Ly-9-like), *ENSOARG00000006215*, *ENA000000008981*, *ENSOARG00000009129*, *ENSOARG00000011375* (blood vessel epicardial substance [*BVES*]), *ENSOARG00000015755*, *ENSOARG00000020354* (novel gene; 53% orthology with bovine monoacylglycerol acyltransferase [*MOGAT1*]), *ENSOARG00000025005*, *ENSOARG00000026030* (novel gene), *GPM6A* (glycoprotein M6A), *RAG1* (recombination activating 1), *STX8* (syntaxin 8). The male-specific thymus set was comprised of *ENSOARG00000007267* (novel gene; t-cell surface glycoprotein *CD1a*-like), *ENSOARG00000016841* (novel gene; 98% orthology with bovine ATP synthase membrane subunit G [*ATP5MG*]), *ENSOARG00000007603*, *SNX25* (sorting nexin 25) and *LDHA* (L-lactate dehydrogenase A chain) **Figure 5D**.

In summary, very few ASE genes were shared across all sheep, and the majority of ASE profiles were private to each sheep. Sex-specific ASE signatures were also detected, but due to the small sample size (n = 3) in both cases, these should be interpreted with caution.

ASE in Stimulated and Unstimulated BMDMs (0 h vs 7 h +LPS)

We examined inducible ASE after 7 h of exposure to LPS in BMDMs using the ICD mode of GeneiASE (Edsgärd et al., 2016). A comparison of LPS-inducible ICD-ASE genes and the genes with background static ASE at 0 and 7 h timepoints, was also performed. We first assessed whether differences between 0 and 7 h could be observed using analysis of static ASE. Individual-specific ASE profiles and a limited number of shared ASE genes were also observed in BMDMs. The total number of genes with static ASE in the BMDMs is shown in **Table 2**.

Shared static ASE across both timepoints and independent of LPS induction was only observed in five genes. These genes have a macrophage associated function and include *ITGB2* (Yee and Hamerman, 2013), *SAA3* (*ENSOARG00000009963*) (Larson et al., 2003; Deguchi et al., 2013), *CD200R1* (*ENSOARG00000019357*) (Ocaña-Guzman et al., 2018), *DCTN5* (*ENSOARG00000017281*) (Habermann et al., 2001), and *MTIF2* (also seen in the tissue analysis above) (Overman et al., 2003).

The ICD-ASE in BMDMs ± LPS captured fewer ASE genes with significant LPS-inducible ASE between the two timepoints than the static analysis of ASE. Moreover, there were large differences in the number of LPS-inducible ASE genes in each individual sheep, indicating significant individual-specific variation in response to LPS stimulation. BMDM cultured from female 2 showed no LPS-inducible response in comparison to male 3 which was a hyper responder with significant inducible ICD-ASE in 28 genes (including 634 informative SNVs total). A detailed breakdown of SNVs, aggregated within each gene, with significant ICD ASE has been summarized in **Figure 6**.

In summary, the ICD-ASE mode of GeneiASE's model was not capable of capturing a complete picture of differential ASE in the BMDM experiment. Static ASE was present in both timepoints; however, there were no inducible ASE genes that were shared across all six sheep. The highly diverse ASE profile of BMDMs was very individual-specific, similar to patterns observed in tissues (**Supplementary Table S1**). These individual-specific differences could be due to individual variation in the innate immune response or experimental variation introduced during primary cell culture or stimulation with LPS.

Condition-Dependent ASE at the SNV Level in BMDMs

To further investigate allelic imbalance at the SNV level without the aggregative gene model of GeneiASE, Fisher's exact test was used. The filtered read counts for bi-allelic SNVs shared by all six sheep BMDMs were selected (n = 646 sites). Allelic read counts of each SNV were tested using Fisher's exact test between 0 h and 7 h (2 × 2 table). Overall, the six sheep shared 646 SNVs with identical allelic genotypes in both timepoints of the BMDM RNA-Seq data set. These SNVs were tested for association with LPS treatment and only four SNVs showed a strong association (FDR < 1×10^{-8}) and 12 SNVs had an FDR between 1×10^{-2} and 1×10^{-8} (**Figure 7**). The highest F-statistic was at rs430667535 Chr17 Pos:50485358T > C, a synonymous variant in ubiquitin C (*UBC*), a polyubiquitin precursor, and also an intronic variant T > C or A > G in pro-apoptotic *BRI3* binding protein (*BRI3BP*). This variant was shown to have a minor allele frequency (MAF) of 0.25 in Texel sheep based on the ISCG annotation [ISGC – Ensembl v.92] (Kijas et al., 2012). The next highest peak was observed on Chr21 under SNVs within *SAA3* gene boundaries (*ENSOARG00000009963*)

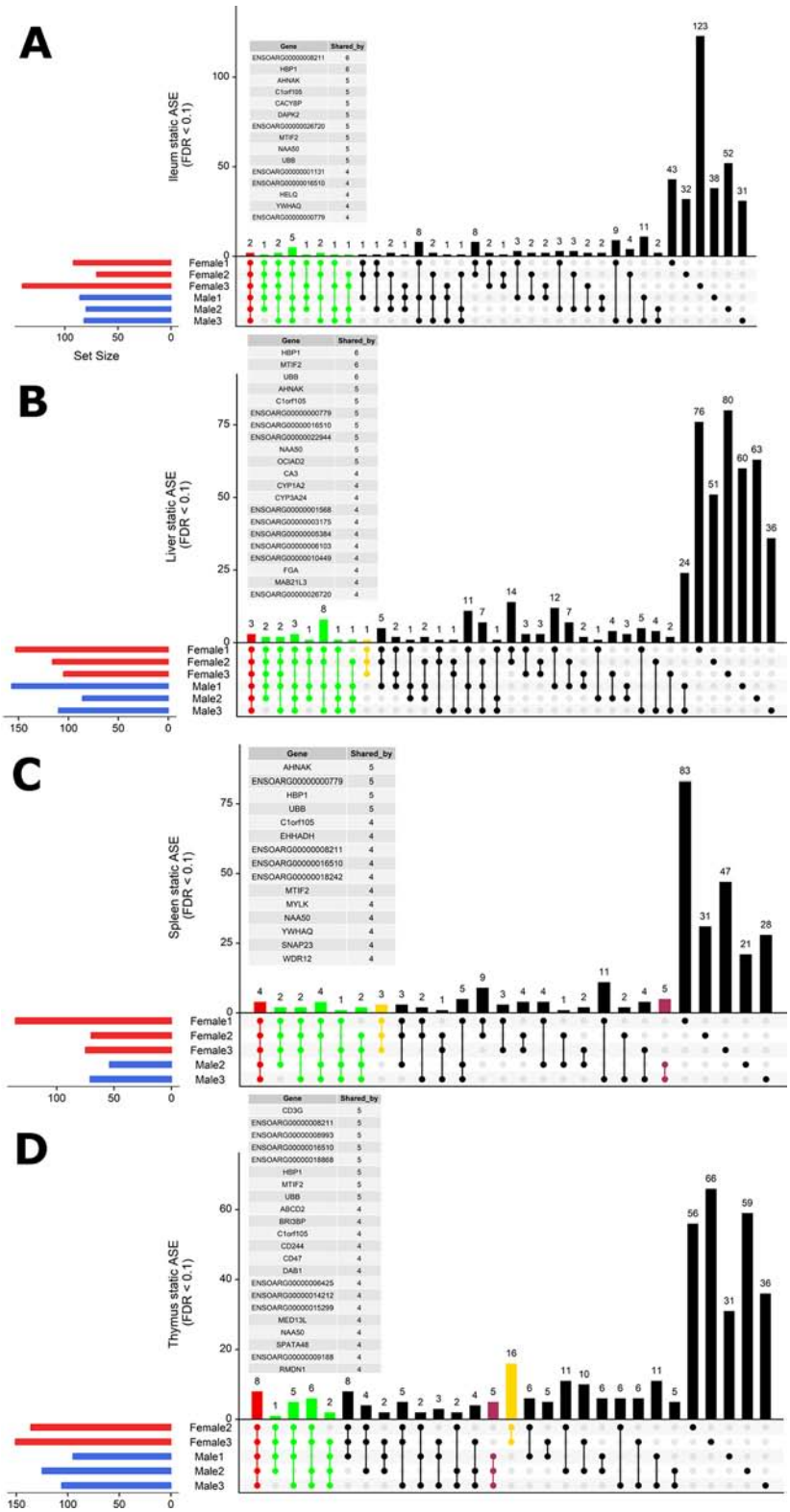


FIGURE 5 | Intersectionality analysis of genes expressing significant ASE across all six sheep. In each tissue from left to right, the set count of genes (dots connected by lines) illustrates the number of sheep sharing the gene. The private sets of genes are located at the far right of each graph (single dots with no line). The intersections are colored in to illustrate the size of the set of shared genes (red [common to all six sheep], green [shared by five or four sheep], yellow [only in females] and purple [only in males]). Detailed lists of genes with ASE shared by at least four sheep are presented above each graph for **(A)** ileum, **(B)** liver, **(C)** spleen, and **(D)** thymus. Two sex-specific sets of genes are highlighted: 16 genes showing ASE only in females (in yellow) and five genes only in males (in purple).

TABLE 2 | Total number of genes with significant static ASE in BMDMs ± LPS.

Sheep	BMDM 0 h –LPS	BMDM 7 h +LPS
Female 1	237	252
Female 2	192	233
Female 3	227	263
Female avg.	219	249
Male 1	205	193
Male 2	300	260
Male 3	260	261
Male avg.	255	238

at the following coordinates: highest FDR peak was observed at SNV rs412192652 (Pos:25826978A > G missense variant [Asn145Asp], FDR = 2.3×10^{-15}) surrounded by rs403064928 (Pos:25826884C > A missense variant [Asp113Glu] in exon 4 of *SAA3* FDR = 3.6×10^{-7}), rs426609498 (Pos:25826845A > G synonymous, FDR = 5.5×10^{-7}) and rs405439099 (Pos:25826990G > C 3'UTR variant, MAF 0.4 in Texel sheep, FDR = 4.1×10^{-5}). This region contains a strong LD block previously reported by the ISGC COMPOSITE population (Ensembl v.92) (**Supplementary Figure S4**), e.g., rs412192652 and rs405439099 pairwise D' statistics = 1. Two further peaks on Chr3 were observed for rs159926581 (Pos:214731375T > C synonymous, FDR = 3.3×10^{-9}) in ribosomal protein L3 (*ENSOARG00000016495*) and rs159822214 (Pos:112164732G > A 3'UTR variant, FDR = 9.5×10^{-3}) in oxysterol binding protein like 8 (*OSBPL8*). The last inducible ASE associated signal was on Chr16 rs420037698 (Pos:6887423G > A, FDR = 6.9×10^{-9}) in *ENSOARG00000004700*, a known synonymous variant in the Texel population (MAF = 0.55). The SNVs and corresponding genes from Fisher's exact test are summarized in **Table 3**.

The LPS-inducible ASE analysis, of SNVs, using Fisher's exact test revealed a different picture not captured by the gene level analysis with the GeneiASE model that aggregates SNVs within each gene (**Figure 6** vs **Figure 7**). The aggregative gene model did not capture any shared ASE genes in the ICD-ASE mode. Although the six sheep shared 646 SNVs and showed highly significant association with stimulation with the LPS (Fisher's exact method), the aggregation of ASE effect size from SNV to gene level (ICD-ASE mode) only detected individual-specific sets of ASE genes in each sheep. This contradicted the results from Fisher's exact test which detected four highly significant LPS-inducible shared regions (FDR < 0.01, 16 SNVs) on chromosomes 3, 16, 17, and 21. For example, the allelic imbalance in the *SAA3* genomic coordinates on chromosome 21 was not detectable in the ICD-ASE model but it was captured by the Fisher's exact test in all individuals (**Figure 7B** chromosome 21).

Fisher's exact analysis at SNV level revealed ASE in response to LPS in variants within *CLDN1*, *ANXA3*, *BRI3BP*, *SAA3,* and *MSR1* (**Table 3**). The anti-inflammatory macrophage marker *CLDN1* (Van den Bossche et al., 2012) and acute-phase inflammation resolution marker *ANXA3* (Yamanegi et al., 2018) have been previously reported with distinct macrophage functions. The macrophage scavenger receptor 1 (*MSR1*) has also been shown to be involved in lipid uptake and migration ability of macrophages (Shi et al., 2019). Three noncoding RNAs (*RF00221* [*snoRD43*], *RF00593* [*snoU83B*], and *RF01151* [*snoU82P*] were among genes corresponding to ASE inducible SNVs. These three snoRNAs all overlap with the genomic coordinates of ribosomal protein L3 (*ENSOARG00000016495*/*RPL3*). The *RF00377* [*snoU6-53*] was also among the ASE-positive targets which overlaps the protein-coding gene *CDS2* (CDP-diacylglycerol synthase 2). Using total RNA-Seq (ribosomal RNA

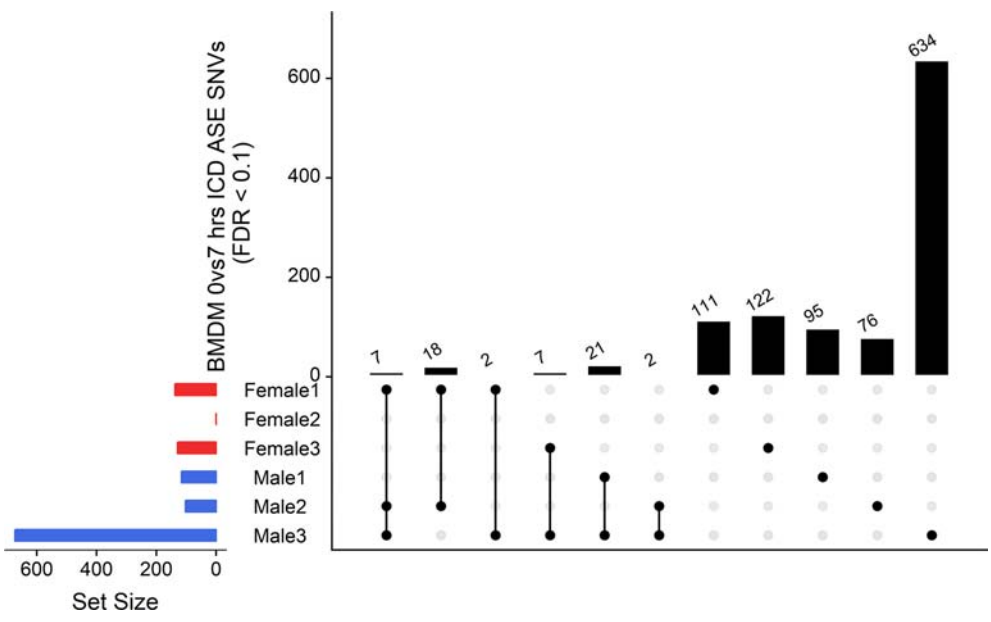


FIGURE 6 | Intersection analysis of SNVs under genes with significant ICD-ASE in the BMDMs ± LPS. From left to right, the set number of genes (dots connected by lines) has been illustrated in order according to the number of sheep sharing the SNV. The private sets of SNVs are located at the far right of each graph (single dots with no line).

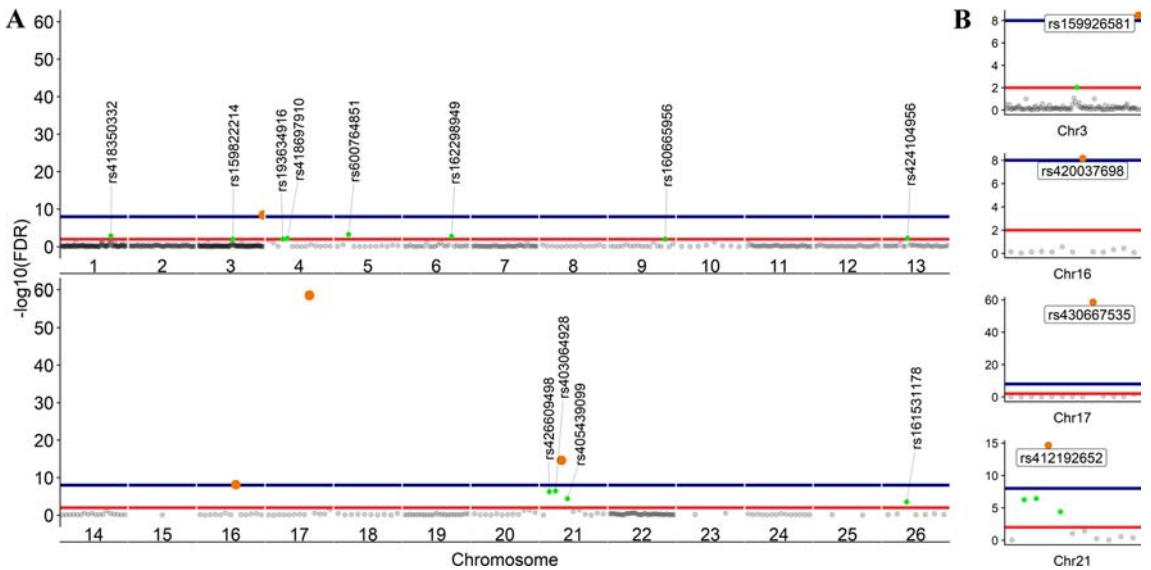


FIGURE 7 | Scatter plot of the adjusted *p* values from Fisher's exact test (unified using Stouffer unification) in BMDMs comparing expression from different alleles at 0 vs 7 h at SNV level (LPS-inducible ASE). **(A)** The graph shows 646 loci exhibiting LPS-inducible allelic imbalance shared across all six sheep. **(B)** Four loci on chromosomes 3, 16, 17, and 21 with false discovery rate (FDR) < 1×10^{-8}. FDR < 1×10^{-2} red line (n = 16 SNVs) and FDR < 1×10^{-8} blue line (n = 4 SNVs).

TABLE 3 | The variant IDs of LPS-inducible ASE SNVs (Fisher's exact) and their respective genes. Data were obtained using Ensembl BioMart query builder. Highly significant SNVs are highlighted in bold [false discovery rate (FDR), < 1×10^{-8}].

ID	CHR	POSITION	GENE ID	GENE NAME
rs418350332	1	195297663	ENSOARG00000020472	CLDN1
rs159822214	3	112164732	ENSOARG00000014876	OSBPL8
rs159926581	**3**	214731375	**ENSOARG00000016495**	–
rs159926581	**3**	214731375	**ENSOARG00000022372**	RF00221
rs159926581	**3**	214731375	**ENSOARG00000024737**	RF00593
rs159926581	**3**	214731375	**ENSOARG00000025124**	RF01151
rs159926581	**3**	214731375	**ENSOARG00000025150**	RF00593
rs193634916	4	53430934	ENSOARG00000001754	MDFIC
rs418697910	4	53430968	ENSOARG00000001754	MDFIC
rs162298949	6	92979793	ENSOARG00000018710	ANXA3
rs162298949	6	92979793	ENSOARG00000018710	ANXA3
rs160665956	9	76759747	ENSOARG00000001261	SPAG1
rs424104956	13	46654781	ENSOARG00000017292	CDS2
rs424104956	13	46654781	ENSOARG00000017292	CDS2
rs424104956	13	46654781	ENSOARG00000022618	RF00377
rs420037698	**16**	**6887423**	**ENSOARG00000004700**	–
rs430667535	**17**	**50485358**	**ENSOARG00000017707**	**BRI3BP**
rs430667535	**17**	**50485358**	**ENSOARG00000018177**	**UBC**
rs403064928	21	25826884	ENSOARG00000009963	SAA3
rs405439099	21	25826990	ENSOARG00000009963	SAA3
rs412192652	**21**	**25826978**	**ENSOARG00000009963**	**SAA3**
rs161531178	26	19663270	ENSOARG00000009626	MSR1

depleted), which includes multiple RNA populations, to generate short read illumina data makes it difficult to pinpoint the origin of the ASE signal to a specific RNA population.

In summary, ASE profiles in BMDMs were highly individual-specific at both gene and SNV level. Moreover, Fisher's exact SNV level analysis discovered shared ASE SNVs where the aggregative gene model of ICD-ASE mode did not, indicating for condition-dependent ASE analysis Fisher's exact test is more accurate and robust at SNV level.

DISCUSSION

This study is the first to investigate global allele-specific expression across tissues from sheep using RNA-Seq data. We focused our analysis on immune-related tissues and cell types from six adult crossbred sheep (T×BF) from the sheep gene expression atlas. ASE profiles were highly individual-specific in the six sheep analyzed in this study. We were able to identify tissue-specific sets of ASE genes, as well as LPS-inducible sets in

the BMDM experiment. Tissue-specific signatures of ASE have been previously reported in similar studies in mouse (Castel et al., 2015; Castel et al., 2016), goat (Cao et al., 2019) and cattle (Chamberlain et al., 2015).

Several steps were taken in the cattle study (Chamberlain et al., 2015) to mitigate the ref allele bias, assign parental origin using whole genome sequences and include MAE variants. The SNV filtration was based on the (Hayes and Daetwyler, 2019) 1000 bull genomes project to confirm the heterozygote sites. In our pipeline, the Ensembl VCF track was used for that purpose. Chamberlain et al. (2015) use a 0.9 allele frequency cutoff (based on read counts) to define and include MAE, and as such have a 1 and 0 inflated allelic ratios. In our pipeline, no allelic ratio cutoff is introduced for inclusion as it is difficult to distinguish between sequencing error and MAE. The minimum read (bi-allelic expression) filtration criteria was applied to exclude highly sequenced loci (Either count/Total >1%) or sequencing errors presenting as rare alleles (min either allele count ≥ 3) which consequentially excludes actual MAE as well as spurious allelic counts. Chamberlain et al. tested 5317 genes (14,495 SNVs) in spleen and detected 382 ASE genes (with min > 1 SNV per gene, similar to this study). Although direct comparison would not be appropriate (because we have excluded MAE variants in our analysis), our analysis of sheep spleen revealed ASE in 86 genes (averaged in five sheep) from 8272 filtered genes (averaged in five sheep). Similarly, in the thymus, the cattle study showed 182 ASE genes from 986 informative genes (9781 SNVs), whereas from 7961 filtered genes in sheep thymus, 134 ASE genes were captured. The differences in the numbers of genes exhibiting ASE between the two studies are likely to be a consequence of the filtration criteria applied, the exclusion of MAE, and species-specific differences between sheep and cow. Results from a more recent study in goat (Cao et al., 2019) more closely reflect our findings. They apply similar filtration criteria to our workflow and discovered 144 ASE genes in liver in comparison to 123 in our sheep liver sets (averaged across six sheep). Other recent studies, including those focusing on production relevant tissues, such as muscle (Guillocheau et al., 2019), have also applied similar stringency in filtration criteria. The filtering criteria we have used for this analysis is stringent and focused on detecting variants of moderate to extreme effects. Further analysis of the data set reducing these criteria might discover additional variants exhibiting ASE across individuals and tissues, but it would also increase the potential risk of false-positive discovery.

For this analysis, we have adapted an ASE analysis workflow with a primary focus on mapping bias removal prior to allele-specific analysis of the transcriptome. The collection of scripts for WASP, used for this analysis, or modified versions of them have been utilized by others for mapping bias removal in reference-guided genomic data sets, e.g., RNA-Seq (Mozaffari et al., 2018; Zhou et al., 2018), Chip-Seq (Pelikan et al., 2018), and for methylomic and epigenetic analysis (Richard Albert et al., 2018).

The ASE analysis pipeline we have adapted for sheep for this study is also adaptable to other species and tissue types with available RNA-Seq data sets. It could be applied, for example, to profile allele-specific expression in the RNA-Seq data sets from livestock species listed on the FAANG data portal (Andersson et al., 2015; Harrison et al., 2018). We used the Ensembl VCF track to capture information at heterozygote loci; however, the individual VCF file from each sheep could also be used in ASE analysis. The latter strategy might enable the capture of rare variants not included in the publicly available VCF tracks but would also raise the issue of normalization/standardization between VCF call sets. The usage of either of these methods will be limited to the number of loci shared by coordinate and bi-allelic genotype (i.e., pervasive ASE discovery). Other studies have compared variants at the RNA and DNA level from the same individual and then removed the DNA variants not present in the RNA-Seq data from the analysis (Guillocheau et al., 2019). We believe that the strength of the pipeline we present is that it does not require parental genotypes and can therefore be applied to other RNA-Seq data sets for livestock where this information is not available.

In our analysis we have not considered either parent-of-origin or breed-of-origin-specific effects in this analysis. For parent-of-origin or breed-of-origin assignment of these ASE profiles, DNA level genotypes from the parents of the six sheep from the gene expression atlas (i.e., Texel sire and Scottish Blackface dam) would be required, and these are unfortunately not available. In this study, ASE expression profiles also might be affected by the direction of the cross (i.e., Texel sire × Scottish Blackface dam). To fully characterize parent-of-origin or breed-of-origin, reciprocal cross experiments would be required. Reciprocal cross studies in mouse (Huang et al., 2017), chicken (Zhuo et al., 2017), and crossbred cattle (Chen et al., 2016b) have shed light on the complexity of such pervasive ASE markers and parent-of-origin effects. Though potentially very interesting, these experiments are lengthy and costly to perform in sheep. Particularly, in this case, because the reciprocal cross (Scottish Blackface dam × Texel sire) is rarely used in the UK sheep industry and as a consequence has limited relevance to production.

Our approach also excludes mono-allelic expression. The minimum filtration criteria utilized in our workflow along with the reference mapping bias removal step ensures an unbiased ASE discovery in the transcriptome by excluding the ambiguity surrounding MAE variants. This form of analysis is based on the principle that absence of evidence (reads) for either allele of a heterozygote site does not directly amount to evidence of their absence, i.e., MAE. The pattern of ASE (ratio of Alt/Ref+Alt) is dependent on the bi-allelic expression of loci within the genomic coordinates of the gene or genomic element of interest. For an ASE effect to be captured by the GeneiASE model, the following criteria must be met: (i) biallelic expression of the locus; (ii) min depth criteria for each allele (min 3 reads, total 10 reads at that site and > 1% of total reads containing that allele); (iii) the allelic imbalance or departure from bi-allelic balanced expression being inducible by an environmental trigger (i.e., LPS in ICD ASE experiment with BMDM data). These stringent criteria secure robust transcriptome-wide ASE discovery while maximizing the usage of read counts from short read RNA-Seq data sets without considering mono-allelic sites. MAE patterns are impossible to differentiate from sequencing error or random nonsense-mediated decay in total RNA-Seq, unless arbitrary cutoffs are introduced, such as ratio of allelic read counts > 0.9 (Chamberlain

et al., 2015) or > 0.7 (Cao et al., 2019). We decided to exclude MAE from this study using read count bi-allelic expression filtration because it is difficult to distinguish between sequencing error and MAE. We do appreciate that this form of filtration might lead to a reduced number of ASE genes discoveries overall and will exclude potentially imprinted loci altogether.

The tissues utilized for ASE analysis in this study (thymus, spleen, liver, and ileum) are highly influential on performance of the immune system. ASE profiles shared across tissues and cell types were limited and instead they tended to be highly specific. We identified tissue-specific ASE in several genes in the thymus, for example, those that are involved in the T cell–mediated immune response, including *CD47* and *CD244*. These tissue-specific and cell-type–specific ASE profiles may underlie the expression of economically important traits, such as disease resistance. Assessment of the connection between economically relevant phenotypes and tissue-specific ASE profiles could be useful for the improvement of genomics enabled sheep breeding programs, particularly those using specialized sire and dam lines (Georges et al., 2018). Loci exhibiting ASE have been associated with production traits including milk-fat percentage (Hayes et al., 2010; Suárez-Vega et al., 2017), trypanotolerance in small ruminants (Kadarmideen et al., 2011; Álvarez et al., 2016), mastitis in goat (Ilie et al., 2018), Johne's disease in cattle (Mallikarjunappa et al., 2018), and Marek's disease in chicken (Maceachern et al., 2011; Meydan et al., 2011; Cheng et al., 2015). Although there is no general consensus currently on the correlation of allelic expression haplotypes and phenotypes under selection in sheep, this form of ASE analysis could pave the way for functional validation at population level (e.g., breed or haplotype-specific aseQTL studies in a larger population of sheep). Examples of population level aseQTL, eQTL, and sQTL (QTLs associated with RNA splicing) already exist for cattle (Wang et al., 2018; Xiang et al., 2018). Knowledge of favorable ASE in critical genes involved in traits of interest could be used as a performance indicator or included as weighted SNVs in genomic prediction algorithms to enhance livestock breeding programs (Georges et al., 2018). Currently, the UK sheep industry is on the cusp of applying genomic prediction, but suitable genomics enabled breeding programs for sheep already exist in New Zealand and Australia (Daetwyler et al., 2010).

CONCLUSIONS

In this study, we characterize extreme to moderate allele-specific expression, at the gene and SNV level, in immune-related tissues and cells from six adult sheep (T×BF) from the sheep gene expression atlas data set. Reference mapping bias removal was an integral component of the analysis pipeline applied in this study. The correction of reference bias prior to obtaining the allelic read counts is a critical step toward true ASE discovery. The workflow developed as part of this manuscript provides an RNA-Seq-only–dependent tool, without the need for individual DNA sequences. We note that the stringent filtering process applied would remove loci where the allelic imbalance was less extreme but might still be of biological significance.

This study is a novel analysis of an existing large-scale complex RNA-Seq data set from sheep. Using the pipeline, we have adapted for this analysis, we were able to identify ASE profiles that were pervasive in each sheep and specific to the tissues and cell types investigated. These tissue and cell type-specific ASE profiles may underlie the expression of economically important traits and could be used to identify variants that could be weighted in genomic prediction algorithms for the improvement of sheep breeding programs. In summary, we have adapted a robust methodology for ASE profiling, using the sheep gene expression atlas data set, and provided a foundation for identifying the regulatory and expressed elements of the genome that are driving complex traits in livestock.

DATA AVAILABILITY

The RNA-Seq sequence data are available *via* the European Nucleotide Archive (ENA) under PRJEB19199 (https://www.ebi.ac.uk/ena/data/view/PRJEB19199). The BAM files were already produced as part of the sheep gene expression atlas following the publicly available protocol (FAANG, 2018) and as described in (Clark et al., 2017). The BAM files have been uploaded to ENA under accession numbers ERZ827944, ERZ827949, ERZ827951, ERZ827955, ERZ827972, ERZ827988, ERZ827995, ERZ827997, ERZ828001, ERZ828016, ERZ828019, ERZ828036, ERZ828044, ERZ828046, ERZ828050, ERZ828070, ERZ828073, ERZ828160, ERZ828167, ERZ828168, ERZ828172, ERZ828188, ERZ828192, ERZ828209, ERZ828215, ERZ828217, ERZ828221, ERZ828240, ERZ828244, ERZ828261, ERZ828268, ERZ828270, ERZ828274, ERZ828293, and ERZ828297. The Oar v3.1 reference FASTA and VCF file from Ensembl v92 were used throughout the pipeline. The ASE analysis pipeline (https://msalavat@bitbucket.org/msalavat/asewrap_public.git) was wrapped using bash scripting on Edinburgh Compute and Data Facility computing resource Eddie Mark 3 (Edinburgh, 2018). All the raw ASE genes data produced by GeneiASE are included in **Supplementary File 2** (Supplementary_file2.zip). The comparison of ASE positive SNVs between imbalance towards the ref or alt alleles is detailed in **Supplementary Figure S7**. The clustering behavior of ASE profiles in BMDMs experiment were explored using Kmeans clustering and Principle Component Analysis (PCA) as shown in **Supplementary Figure S8** and **S9** respectively. All the supplementary material are available at (https://doi.org/10.6084/m9.figshare.8035799.v1).

ETHICS STATEMENT

Approval was obtained from The Roslin Institute, University of Edinburgh's Animal Work and Ethics Review Body (AWERB). All animal work was carried out under the regulations of the Animals (Scientific Procedures) Act 1986.

AUTHOR CONTRIBUTIONS

MS and EC coordinated and designed the analysis component of the study with assistance from SB and SP-V. MS and SP-V

designed, optimized and tested the ASE pipeline. DH acquired the funding for the sheep gene expression atlas project. MM, EC, and DH designed the LPS experiment and generated the data. MM performed the LPS stimulation of bone marrow derived macrophages and RNA extraction. SB performed all bioinformatic analyses prior to analysis with the ASE pipeline. MS performed ASE analysis, visualization of the results and wrote the manuscript. EC contributed to manuscript editing and drafting. All authors read and approved the final manuscript.

FUNDING

The work was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) (http://www.bbsrc.ac.uk) Grant BB/L001209/1 ('Functional Annotation of the Sheep Genome'). Also, BBSRC Institute Strategic Program Grants: 'Farm Animal Genomics' (BBS/E/D/20211550), 'Transcriptomes, Networks and Systems' (BBS/E/D/20211552), 'Improving Animal Production and Welfare' (BB/P013759/1) and 'Blue Prints for Healthy Animals' (BB/P013732/1). Edinburgh Genomics is partly supported through core grants from BBSRC (BB/J004243/1), NERC (http://www.nerc.ac.uk) (R8/H10/56) and the Medical Research Council (MRC) (https://www.mrc.ac.uk) (MR/K001744/1). SB was supported by the Roslin Foundation. EC is supported by the University of Edinburgh Chancellor's Fellowship programme. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank University of Edinburgh Computing and Data Support staff (viz. Dr Andy Law and Dr Steve Thorn) for their input toward optimization of the pipeline and efficient parallel computing. Rachel Young and Lucas Lefevre isolated the bone marrow derived macrophages for the sheep gene expression atlas project and Iseabail Farquhar helped to collect and archive the tissue samples. The authors would also like to thank Professor Kim Summers and Professor Mick Watson for their comments, which helped to improve the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://doi.org/10.6084/m9.figshare.8035799.v1

REFERENCES

Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I., and Goyache, F. (2016). Lack of specific alleles for the bovine chemokine (C-X-C) receptor type 4 (CXCR4) gene in West African cattle questions its role as a candidate for trypanotolerance. *Infect. Genet. Evol.* 42, 30–33. doi: 10.1016/j.meegid.2016.04.029

Andersson, L., Archibald, A. L., Bottema, C. D., Brauning, R., Burgess, S. C., Burt, D. W., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16, 57. doi: 10.1186/s13059-015-0622-4

Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a018382. doi: 10.1101/cshperspect.a018382

Barton, S. J., Crozier, S. R., Lillycrop, K. A., Godfrey, K. M., and Inskip, H. M. (2013). Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. *BMC Genomics* 14, 161. doi: 10.1186/1471-2164-14-161

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing on JSTOR. *Source J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Bush, S. J., McCulloch, M. E. B., Muriuki, C., Salavati, M., Davis, G. M., Farquhar, I. L., et al. (2019). Comprehensive Transcriptional Profiling of the Gastrointestinal Tract of Ruminants from Birth to Adulthood Reveals Strong Developmental Stage Specific Gene Expression. *G3 (Bethesda)* 9, 359–373. doi: 10.1534/g3.118.200810.

Cao, Y., Xu, H., Li, R., Gao, S., Chen, N., Luo, J., et al. (2019). Genetic basis of phenotypic differences between Chinese Yunling black goats and Nubian goats revealed by allele-specific expression in their F1 Hybrids. *Front. Genet.* 10, 145. doi: 10.3389/fgene.2019.00145

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195. doi: 10.1186/s13059-015-0762-6

Castel, S. E., Mohammadi, P., Chung, W. K., Shen, Y., and Lappalainen, T. (2016). ARTICLE Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat. Commun.* 7, 12817. doi: 10.1038/ncomms12817

Chamberlain, A. J., Vander Jagt, C. J., Hayes, B. J., Khansefid, M., Marett, L. C., Millen, C. A., et al. (2015). Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* 16, 993. doi: 10.1186/s12864-015-2174-0

Chen, J., Rozowsky, J., Galeev, T. R., Harmanci, A., Kitchen, R., Bedford, J., et al. (2016a). A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat. Commun.* 7, 11101. doi:10.1038/ncomms11101

Chen, Z., Hagen, D. E., Wang, J., Elsik, C. G., Ji, T., Siqueira, L. G., et al. (2016b). Global assessment of imprinted gene expression in the bovine conceptus by next generation sequencing. *Epigenetics* 11, 501–516. doi: 10.1080/15592294.2016.1184805

Cheng, H. H., Perumbakkam, S., Pyrkosz, A. B., Dunn, J. R., Legarra, A., and Muir, W. M. (2015). Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. *BMC Genomics* 16, 816. doi: 10.1186/s12864-015-2016-0

Clark, E. L., Bush, S. J., McCulloch, M. E. B., Farquhar, I. L., Young, R., Lefevre, L., et al. (2017). A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLoS Genet.* 13, e1006997. doi: 10.1371/journal.pgen.1006997

Daetwyler, H. D., Hickey, J. M., Henshall, J. M., Dominik, S., Gredler, B., Van Der Werf, J. H. J., et al. (2010). Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim. Prod. Sci.* 50, 1004–1010. doi: 10.1071/AN10096

Degner, J. F., Marioni, J. C., Pai, A. A., Pickrell, J. K., Nkadori, E., Gilad, Y., et al. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinforma. Orig. Pap.* 25, 3207–321210. doi: 10.1093/bioinformatics/btp579

Deguchi, A., Tomita, T., Omori, T., Komatsu, A., Ohto, U., Takahashi, S., et al. (2013). Serum amyloid A3 binds MD-2 to activate p38 and NF-κB pathways in a MyD88-dependent manner. *J. Immunol.* 191, 1856–1864. doi: 10.4049/jimmunol.1201996

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* 8, e1002600. doi: 10.1371/journal.pgen.1002600

Dewey, M. (2016). Comparison of various methods for meta–analysis of p–values. http://www.dewey.myzen.co.uk/home.html, 1–8. Available at: http://www.dewey.myzen.co.uk/meta/comparemet.pdf [Accessed May 8, 2018].

Dewey, M. (2019). {metap}: meta-analysis of significance values. R package version 1.1. 1–26. Available at: http://www.dewey.myzen.co.uk/meta/meta.html.

Edsgärd, D., Iglesias, M. J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J., et al. (2016). GeneiASE: detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci. Rep.* 6, 21134. doi: 10.1038/srep21134

FAANG. (2018). Sheep Atlas RNA-Seq Processing Pipeline SOP. Available at: ftp://ftp.faang.ebi.ac.uk/ftp/protocols/analysis/Roslin_SOP_Alignment-based RNA-Seq-Processing_20170917.pdf

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Res.* 46, D649–D655. doi: 10.1093/nar/gkx1132

García-Gámez, E., Reverter, A., Whan, V., McWilliam, S. M., Arranz, J. J., Kijas, J., et al. (2011). Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in Merino Sheep. *PLoS One* 6, e21158. doi: 10.1371/journal.pone.0021158

Georges, M., Charlier, C., and Hayes, B. (2018). Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20, 135–156. doi: 10.1038/s41576-018-0082-2

Georgoudaki, A.-M., Khodabandeh, S., Puiac, S., Persson, C. M., Larsson, M. K., Lind, M., et al. (2015). CD244 is expressed on dendritic cells and regulates their functions. *Immunol. Cell Biol.* 93, 581–590. doi: 10.1038/icb.2014.124

Goodhead, I., Archibald, A., Amwayi, P., Brass, A., Gibson, J., Hall, N., et al. (2010). A comprehensive genetic analysis of candidate genes regulating response to *Trypanosoma congolense* infection in mice. *PLoS Negl. Trop. Dis.* 4, e880. doi: 10.1371/journal.pntd.0000880

GTEx Consortium, Gte., Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277

Guillocheau, G. M., El Hou, A., Meersseman, C., Esquerré, D., Rebours, E., Letaief, R., et al. (2019). Survey of allele specific expression in bovine muscle. *Sci. Rep.* 9, 4297. doi: 10.1038/s41598-019-40781-6

Gutierrez-Arcelus, M., Baglaenko, Y., Arora, J., Hannes, S., Luo, Y., Amariuta, T., et al. (2019). Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *bioRxiv* 599449. doi: 10.1101/599449

Habermann, A., Schroer, T. A., Griffiths, G., and Burkhardt, J. K. (2001). Immunolocalization of cytoplasmic dynein and dynactin subunits in cultured macrophages: enrichment on early endocytic organelles. *J. Cell Sci.* 114, 229–240.

Hach, F., Sarrafi, I., Hormozdiari, F., Alkan, C., Eichler, E. E., and Sahinalp, S. C. (2014). mrsFAST-Ultra: a compact, SNP-aware mapper for high performance sequencing applications. *Nucleic Acids Res.* 42, W494–W500. doi: 10.1093/nar/gku370

Harrison, P. W., Fan, J., Richardson, D., Clarke, L., Zerbino, D., Cochrane, G., et al. (2018). FAANG, establishing metadata standards, validation and best practices for the farmed and companion animal community. *Anim. Genet.* 49, 520–526. doi: 10.1111/age.12736

Hasin-Brumshtein, Y., Hormozdiari, F., Martin, L., van Nas, A., Eskin, E., Lusis, A. J., et al. (2014). Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics* 15, 471. doi: 10.1186/1471-2164-15-471

Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annu. Rev. Anim. Biosci.* 7, 89–102. doi: 10.1146/annurev-animal-020518-115024

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., and Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* 6, e1001139. doi: 10.1371/journal.pgen.1001139

Hodgkinson, A., Grenier, J.-C., Gbeha, E., and Awadalla, P. (2016). A haplotype-based normalization technique for the analysis and detection of allele specific expression. *BMC Bioinformatics*. 17, 364. doi: 10.1186/s12859-016-1238-8

Huang, W. C., Ferris, E., Cheng, T., Hörndli, C. S., Gleason, K., Tamminga, C., et al. (2017). Diverse non-genetic, allele-specific expression effects shape genetic architecture at the cellular level in the mammalian brain. *Neuron* 93, 1094–1109.e7. doi: 10.1016/j.neuron.2017.01.033

Ilie, D. E., Kusza, S., Sauer, M., and Gavojdian, D. (2018). Genetic characterization of indigenous goat breeds in Romania and Hungary with a special focus on genetic resistance to mastitis and gastrointestinal parasitism based on 40 SNPs. *PLoS One* 13, e0197051. doi: 10.1371/journal.pone.0197051

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173. doi: 10.1126/science.1252806

Kadarmideen, H. N., Watson-Haigh, N. S., and Andronicos, N. M. (2011). Systems biology of ovine intestinal parasite resistance: disease gene modules and biomarkers. *Mol. BioSyst.* 7, 235–246. doi: 10.1039/C0MB00190B

Kaul, S., Mittal, S. K., Feigenbaum, L., Kruhlak, M. J., and Roche, P. A. (2015). Expression of the SNARE protein SNAP-23 is essential for cell survival. *PLoS One* 10, e0118311. doi: 10.1371/journal.pone.0118311

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10, e1001258. doi: 10.1371/journal.pbio.1001258

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Larson, M. A., Wei, S. H., Weber, A. T. A., Weber, A. T. A., and McDonald, T. L. (2003). Induction of human mammary-associated serum amyloid A3 expression by prolactin or lipopolysaccharide. *Biochem. Biophys. Res. Commun.* 301, 1030–1037. doi: 10.1016/S0006-291X(03)00045-7

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352.

Maceachern, S., Muir, W. M., Crosby, S., and Cheng, H. H. (2011). Genome-wide identification of allele-specific expression (ASE) in response to Marek's disease virus infection using next generation sequencing. *BMC Proc.* 5, S14. doi: 10.1186/1753-6561-5-S4-S14

Mallikarjunappa, S., Sargolzaei, M., Brito, L. F., Meade, K. G., Karrow, N. A., and Pant, S. D. (2018). Short communication: uncovering quantitative trait loci associated with resistance to *Mycobacterium avium* ssp. paratuberculosis infection in Holstein cattle using a high-density single nucleotide polymorphism panel. *J. Dairy Sci.* 101, 7280–7286. doi: 10.3168/jds.2018-14388

Mayba, O., Gilbert, H. N., Liu, J., Haverty, P. M., Jhunjhunwala, S., Jiang, Z., et al. (2014). MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome Biol.* 15, 405. doi: 10.1186/s13059-014-0405-3

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

McNerney, M. E., Lee, K.-M., and Kumar, V. (2005). 2B4 (CD244) is a non-MHC binding receptor with multiple functions on natural killer cells and CD8+ T cells. *Mol. Immunol.* 42, 489–494. doi: 10.1016/j.molimm.2004.07.032

Meydan, H., Yildiz, M. A., Dodgson, J. B., and Cheng, H. H. (2011). Allele-specific expression analysis reveals CD79B has a cis-acting regulatory element that responds to Marek's disease virus infection in chickens. *Poult. Sci.* 90, 1206–1211. doi: 10.3382/ps.2010-01295

Miao, Z., Alvarez, M., Pajukanta, P., and Ko, A. (2018). ASElux: an ultra-fast and accurate allelic reads counter. *Bioinformatics* 34, 1313–1320. doi: 10.1093/bioinformatics/btx762

Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., et al. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464, 773–777. doi: 10.1038/nature08903

Mozaffari, S. V., Stein, M. M., Magnaye, K. M., Nicolae, D. L., and Ober, C. (2018). Parent of origin gene expression in a founder population identifies two new candidate imprinted genes at known imprinted regions. *PLoS One* 13, e0203906. doi: 10.1371/journal.pone.0203906

Mueller, O., Lightfoot, S., and Schroeder, A. (2004). RNA Integrity Number (RIN) – Standardization of RNA quality control application. *Agil. Appl. Note, Publ. 5989-1165EN*, 1–8. doi: 10.1101/gr.189621.115.7

Nitta, T., and Takahama, Y. (2007). The lymphocyte guard-IANs: regulation of lymphocyte survival by IAN/GIMAP family proteins. *Trends Immunol.* 28, 58–65. doi: 10.1016/j.it.2006.12.002

Ocaña-Guzman, R., Vázquez-Bolaños, L., and Sada-Ovalle, I. (2018). Receptors that inhibit macrophage activation: mechanisms and signals of regulation and tolerance. *J. Immunol. Res.* 2018, 8695157. doi: 10.1155/2018/8695157

Overman, R. G., Enderle, P. J., Farrow, J. M., Wiley, J. E., and Farwell, M. A. (2003). The human mitochondrial translation initiation factor 2 gene (MTIF2): Transcriptional analysis and identification of a pseudogene. *Biochim. Biophys. Acta Gene Struct. Expr.* 1628, 195–205. doi: 10.1016/S0167-4781(03)00144-1

Pelikan, R. C., Kelly, J. A., Fu, Y., Lareau, C. A., Tessneer, K. L., Wiley, G. B., et al. (2018). Enhancer histone-QTLs are enriched on autoimmune risk haplotypes and influence gene expression within chromatin networks. *Nat. Commun.* 9, 2905. doi: 10.1038/s41467-018-05328-9

Pounds, S., and Cheng, C. (2006). Robust estimation of the false discovery rate. *Bioinformatics* 22, 1979–1987. doi:10.1093/bioinformatics/btl328

R Core Team. (2017). R: A language and environment for statistical computing. *R Found. Stat. Comput*. Available at: https://www.r-project.org/.

Raghupathy, N., Choi, K., Vincent, M. J., Beane, G. L., Sheppard, K. S., Munger, S. C., et al. (2018). Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. *Bioinformatics* 34, 2177–2184. doi: 10.1093/bioinformatics/bty078

Renfree, M. B., Hore, T. A., Shaw, G., Graves, J. A. M., Pask, A. J., Marshall Graves, J. A., et al. (2009). Evolution of genomic imprinting: insights from marsupials and monotremes. *Annu. Rev. Genomics Hum. Genet.* 10, 241–262. doi: 10.1146/annurev-genom-082908-150026

Richard Albert, J., Koike, T., Younesy, H., Thompson, R., Bogutz, A. B., Karimi, M. M., et al. (2018). Development and application of an integrated allele-specific pipeline for methylomic and epigenomic analysis (MEA). *BMC Genomics* 19, 463. doi: 10.1186/s12864-018-4835-2

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., et al. (2011). AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 7, 522. doi: 10.1038/msb.2011.54

Shi, Q., Zhuang, F., Liu, J.-T., Li, N., Chen, Y.-X., Su, X.-B., et al. (2019). Single-cell analyses reveal functional classification of dendritic cells and their potential roles in inflammatory disease. *FASEB J.* 33, 3784–3794. doi: 10.1096/fj.201801489R

Smith, R. M., Webb, A., Papp, A. C., Newman, L. C., Handelman, S. K., Suhy, A., et al. (2013). Whole transcriptome RNA-Seq allelic expression in human brain. *BMC Genomics.* 14, 571. doi: 10.1186/1471-2164-14-571

Stevenson, K. R., Coolon, J. D., and Wittkopp, P. J. (2013). Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC Genomics* 14, 1. doi: 10.1186/1471-2164-14-536

Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G., and Arranz, J. J. (2017). Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC Genomics* 18, 170. doi: 10.1186/s12864-017-3581-1

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J. K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* 12, 1061–1063. doi: 10.1038/nmeth.3582

Van den Bossche, J., Laoui, D., Morias, Y., Movahedi, K., Raes, G., De Baetselier, P., et al. (2012). Claudin-1, Claudin-2 and Claudin-11 genes differentially associate with distinct types of anti-inflammatory macrophages *in vitro* and with parasite- and tumour-elicited macrophages *in vivo. Scand. J. Immunol.* 75, 588–598. doi: 10.1111/j.1365-3083.2012.02689.x

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 11.10.1–33. doi: 10.1002/0471250953.bi1110s43

Wang, M., Hancock, T. P., Chamberlain, A. J., Vander Jagt, C. J., Pryce, J. E., Cocks, B. G., et al. (2018). Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. *BMC Genomics* 19, 395. doi: 10.1186/s12864-018-4800-0

Wickham, H., Francois, R., Henry, L., and Müller, K. (2019). Package 'dplyr'. A grammar of data manipulation. *R Packag. version 0.8.0.1*, 1–88.

Wong, P. P. C., Daneman, N., Volchuk, A., Lassam, N., Wilson, M. C., Klip, A., et al. (1997). Tissue distribution of SNAP-23 and its subcellular localization in 3T3-L1 cells. *Biochem. Biophys. Res. Commun.* 230, 64–68. doi: 10.1006/bbrc.1996.5884

Xiang, R., Hayes, B. J., Vander Jagt, C. J., MacLeod, I. M., Khansefid, M., Bowman, P. J., et al. (2018). Genome variants associated with RNA splicing variation in bovine are extensively shared between tissues Ruidong. *BMC Genomics* 19, 521. doi: 10.1186/s12864-018-4902-8

Xin, H., Lee, D., Hormozdiari, F., Yedkar, S., Mutlu, O., and Alkan, C. (2013). Accelerating read mapping with FastHASH. *BMC Genomics* 14, S13. doi: 10.1186/1471-2164-14-s1-s13

Yamanegi, K., Yamada, N., Nakasho, K., and Nishiura, H. (2018). Participation of delta annexin A3 in the ribosomal protein S19 C-terminus-dependent inhibitory mechanism of the neutrophil C5a receptor through delta lactoferrin. *Pathol. Int.* 68, 109–116. doi: 10.1111/pin.12626

Yee, N. K., and Hamerman, J. A. (2013). β(2) integrins inhibit TLR responses by regulating NF-κB pathway and p38 MAPK activation. *Eur. J. Immunol.* 43, 779–792. doi: 10.1002/eji.201242550

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46, D754–D761. doi: 10.1093/nar/gkx1098

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhou, Z.-Y., Hu, Y., Li, A., Li, Y.-J., Zhao, H., Wang, S.-Q., et al. (2018). Genome wide analyses uncover allele-specific RNA editing in human and mouse. *Nucleic Acids Res.* 46, 8888–8897. doi: 10.1093/nar/gky613

Zhuo, Z., Lamont, S. J., and Abasht, B. (2017). RNA-Seq analyses identify frequent allele specific expression and no evidence of genomic imprinting in specific embryonic tissues of chicken. *Sci. Rep.* 7, 11944. doi: 10.1038/s41598-017-12179-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Salavati, Bush, Palma-Vera, McCulloch, Hume and Clark. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 October 2019
doi: 10.3389/fgene.2019.00902

Genetic Parameters for Yolk Cholesterol and Transcriptional Evidence Indicate a Role of Lipoprotein Lipase in the Cholesterol Metabolism of the Chinese Wenchang Chicken

*Xingyong Chen[1,2], Wenjun Zhu[1], Yeye Du[1], Xue Liu[1] and Zhaoyu Geng[1,2]**

[1] College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [2] Anhui Province Key Laboratory of Local Livestock and Poultry Genetic Resource Conservation and Bio-breeding, Anhui Agricultural University, Hefei, China

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Carl Joseph Schmidt, University of Delaware, United States
Xiangdong Ding, China Agricultural University (CAU), China

***Correspondence:**
Zhaoyu Geng
gzy@ahau.edu.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *03 April 2019*
Accepted: *26 August 2019*
Published: *02 October 2019*

Citation:
Chen X, Zhu W, Du Y, Liu X and Geng Z (2019) Genetic Parameters for Yolk Cholesterol and Transcriptional Evidence Indicate a Role of Lipoprotein Lipase in the Cholesterol Metabolism of the Chinese Wenchang Chicken. Front. Genet. 10:902. doi: 10.3389/fgene.2019.00902

The yolk cholesterol has been reported to affect egg quality and breeding performance in chickens. However, the genetic parameters and molecular mechanisms regulating yolk cholesterol remain largely unknown. Here, we used the Wenchang chicken, a Chinese indigenous breed with a complete pedigree, as an experimental model, and we examined 24 sire families (24 males and 240 females) and their 362 daughters. First, egg quality and yolk cholesterol content were determined in 40-week-old chickens of two consecutive generations, and the heritability of these parameters was analyzed using the half-sib correlation method. Among first-generation individuals, the egg weight, egg shape index, shell strength, shell thickness, yolk weight, egg white height, Haugh unit, and cholesterol content were 45.36 ± 4.44 g, 0.81 ± 0.12, 3.07 ± 0.92 kg/cm^2, 0.340 ± 0.032 mm, 15.57 ± 1.64 g, 3.36 ± 1.15 mm, 58.70 ± 12.33, and 274.3 ± 36.73 mg/egg, respectively. When these indexes were compared to those of the following generation, no statistically significant difference was detected. Although yolk cholesterol content was not associated with egg quality in females, an increase in yolk cholesterol content was correlated with increased yolk weight and albumin height in sire families ($p < 0.05$). Moreover, the heritability estimates for the yolk cholesterol content were 0.328 and 0.530 in female and sire families, respectively. Therefore, the yolk cholesterol content was more strongly associated with the sire family. Next, chickens with low and high yolk cholesterol contents were selected for follicular membrane collection. Total RNA was extracted from these samples and used as a template for transcriptional sequencing. In total, 375 down- and 578 upregulated genes were identified by comparing the RNA sequencing data of chickens with high and low yolk cholesterol contents. Furthermore, Gene Ontology term and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses indicated the involvement of energy metabolism and immune-related pathways in yolk cholesterol deposition. Several genes participating in the regulation of the yolk cholesterol content were located on the sex chromosome Z, among which lipoprotein lipase (*LPL*) was associated with the peroxisome proliferator-activated receptor signaling pathway

and the Gene Ontology term cellular component. Collectively, our data suggested that the ovarian steroidogenesis pathway and the downregulation of *LPL* played critical roles in the regulation of yolk cholesterol content.

Keywords: heritability, lipoprotein lipase, Wenchang chicken, yolk cholesterol, egg quality

INTRODUCTION

On the day of hatch, most of the yolk sac has been absorbed by the bird, which provides sufficient nutrition for the first days (0–3 days) posthatch (Yair and Uni, 2011). Moreover, it is widely accepted that both growth and breeding performance of birds depend largely on their early health (Yadgary et al., 2010). Therefore, egg yolk quality plays an essential role in maintaining early health and later breeding performance. The main components of the egg yolk are triglyceride, cholesterol, lecithin, vitamins, and minerals (Ding et al., 2017). Previous studies have suggested that cholesterol intake from eggs can affect human health, causing dyslipidemia, hyperlipidemia, atherosclerosis, or cardiovascular diseases (Andersen et al., 2013; Omole and Ighodaro, 2013). Nevertheless, yolk cholesterol is essential for egg production and embryo development. Indeed, in hens that had decreased or insufficient cholesterol synthesis to maintain embryonic development, egg production was reduced or stopped (Janjira, 2017). Furthermore, cholesterol homeostasis is essential and correlates with egg hatchability. While hatchability was increased when the yolk cholesterol content reached a certain level, it was decreased when cholesterol levels increased further and exceeded a certain threshold (Dikmen and Sahan, 2007).

Yolk cholesterol is mainly derived from *de novo* synthesis, and only a small portion is supplemented by feeding, which indicates that yolk cholesterol might be affected by both genetic and nutritional factors (Griffin, 1992; Klkin et al., 1997). Previous studies have reported that the yolk cholesterol concentration varied among breeds ranging from 10 to 100 mmol/L with a normal distribution and was positively correlated with embryo mortality during hatching (Panda et al., 2003; Yang et al., 2013). These observations support the notion that genetic factors might regulate yolk cholesterol. Moreover, cholesterol is found at relatively low levels in feeding, which further suggests that yolk cholesterol is mainly affected by the genetic makeup of the bird (Sreenivas et al., 2013). Accordingly, if the heritability of yolk cholesterol is high, then individual selection could be used. However, if the heritability of yolk cholesterol is moderate or low, then sire selection should be preferred.

In mice, oocyte-derived bone morphogenetic protein 15 (BMP15) and growth differentiation factor 9 (GDF9) have been shown to promote cholesterol biosynthesis in cumulus cells as a compensation mechanism for cholesterol production deficiencies in the oocyte (Su et al., 2008). Furthermore, the *cyp19a1*, *cyp17a1*, tesc, *apoc1*, and *star* genes have been reported to play roles in the regulation of steroidogenesis during oocyte maturation in both trout and Xenopus (Gohin et al., 2010). Moreover, feeding hens with a diet supplemented in alfalfa saponin extract has been shown to decrease the yolk cholesterol content. This decrease in yolk cholesterol was associated with increased expression levels of cholesterol 7 alpha-hydroxylase and apolipoprotein H in the liver and decreased expression levels of very low-density lipoprotein (VLDL) receptor, apolipoprotein B, apovitellenin-1, and vitellogenin in the oocyte (Zhou et al., 2014). Nevertheless, little remains known about the molecular mechanisms underlying the regulation of yolk cholesterol in chicken.

In this study, we used as an experimental model a group of Wenchang chickens, an indigenous Chinese breed with a detailed pedigree. Egg quality was determined in two consecutive generations, and genetic parameters were evaluated in individuals and sire families. Moreover, follicular membrane was collected from hens with either low or high yolk cholesterol content, and transcriptional sequencing was used to screen for candidate genes and signal pathways involved in the regulation of cholesterol synthesis.

MATERIALS AND METHODS

Birds Management

All birds used in this study were Wenchang chickens, a Chinese indigenous breed with a complete pedigree. A total of 24 sire families (24 males and 240 females) and 362 daughters (equality distributed among the sire families with pure breeding) were raised with one bird *per* cage and maintained on a 16 L/8 D (16 h light and 8 h dark) photoperiod during egg laying. At 40 weeks of age, eggs and follicular tissues were collected for quality and yolk cholesterol analysis. Hens were artificially inseminated, and all birds were kept at 15–20°C during the egg-laying period. Egg quality and cholesterol content were determined in two successive generations.

All experimental procedures were performed following guidelines developed by the China Council on Animal Care and Protocols and were approved by the Animal Care and Use Committee of Anhui Agricultural University, China (permission no. SYDW-P2017062801).

Egg Quality and Yolk Cholesterol Analysis

Three eggs were collected from each bird within five consecutive days, and egg quality was assessed within 24 h after collection. A digital scale (accuracy: 0.01 g) was used to measure the weight of each egg. An electronic digital caliper was used to measure the longitudinal diameter (LE) and the transverse diameter (WE) of each egg, and the egg shape index was defined as the WE/LE ratio. Shell strength was measured using an eggshell force gauge

(model II, Robotmation, Tokyo, Japan). Then, the egg was broken onto a flat surface, and the height of the inner thick albumen (egg white) was measured using an egg analyzer (model EA-01, ORKA Food Technology, Ramat HaSharon, Israel). The yolk was separated from the albumen, weighed, and stored at -20°C for cholesterol determination. The shell thickness was measured using a digital Vernier caliper (model NFN380, Fujihira Industry, Tokyo, Japan).

After weighing the yolk, ~0.1 g of yolk was transferred to a 1.5-ml tube. Nine times by weight of anhydrous ethanol were added to the yolk, and the mixture was mechanically homogenized for 30 s at 50 Hz in an ice water bath. Next, all samples were centrifuged for 10 min at 2,500 rpm, and 25 μl of the supernatant was transferred into a well of a 96-well plate. After adding 250 μl of working solution (50 mmol/L Good's buffer, 5 mmol/L phenol, 0.3 mmol/L 4-AAP, ≥50 KU/L cholesteryl esterase, ≥25 KU/L cholesterol oxidase, and ≥1.3 KU/L peroxidase) to each well, the solution was mixed and incubated for 10 min. The optical density (OD) was measured at wavelength of 510 nm, and the cholesterol content was calculated using the following formula: cholesterol content (mg) = (sample OD - blank OD)/(corrected OD - blank OD) × dilution factor × yolk weight × 386.6535/1,000.

Follicular Tissue Collection, Total RNA Extraction, and cDNA Library Construction

After yolk cholesterol been determined, birds with the lowest (L group) and highest (H group) yolk cholesterol content were selected for follicular tissue collection. For each group, three hens at 41 weeks of age were killed ~22 h after ovulation, and then, the ovaries were collected rapidly and kept on ice. Three largest (25–30 mm) yellow preovulatory follicles were isolated from each ovary. The yolk was squeezed out, and the granulosa layer was collected, divided into two parts, and immediately stored in liquid nitrogen for RNA isolation.

Total RNA was isolated from individual samples using the OMEGA total RNA extraction kit (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer's recommendations. RNA integrity number and quality were analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, US). Then, qualified total RNA was further purified using an RNase-Free DNase Set (Qiagen, Hilden, Germany). Purified total RNA was used for the construction of a complementary DNA (cDNA) library and subsequent sequencing (NEB Next Ultra Directional RNA Library Prep Kit for Illumina; New England Biolabs, Ipswich, MA, USA). The remaining RNA from each sample was reverse transcribed and stored at -80°C for RNA sequencing (RNA-Seq) results validation *via* real-time quantitative PCR (RT-qPCR).

RNA-Seq

Following messenger RNA purification using Agencourt AMPure XP beads (Beckman, Brea, CA, USA), the first and second cDNA strands were synthesized using the SuperScriptII Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations. Next, double-stranded cDNA was end repaired, adenylated, and ligated to NEBNext Adaptors (New England Biolabs) according to the manufacturer's recommendations. The cDNA fragments of 150–200 bp were selected using the Agencourt AMPure XP system (Beckman), and PCR was performed using the Phusion High-Fidelity DNA polymerase (New England Biolabs), universal PCR primers, and an Index (X) primer. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA) according to the manufacturer's recommendations. After clustering, the libraries were sequenced using a paired-end 2 × 125 bp lane on an Illumina HiSeq 4000 platform (Shanghai Personal Biotechnology, Shanghai, China).

Filtering of Raw Data and Mapping of High-Quality Reads to the Chicken Reference Genome

Six libraries from each group (n = 3) were sequenced. First, raw reads in FASTQ format were filtered to generate clean reads by removing reads containing adapters or ambiguous nucleotides and reads of low quality, as described by Wang et al. (2017a). Then, the filtered reads were mapped to the chicken reference genome (Gallus_gallus-5.0) using the spliced mapping algorithm of Tophat (version 2.0.9) with no more than two mismatches. Basic mapping statistics, mapped reads distribution across the chicken genome, and annotated genes were determined to evaluate the randomness of the distribution.

Calculation of Gene Expression Level

Gene expression level was calculated using the Cufflinks suite (version 2.1.1) on Tophat output. In brief, the specific gene location was obtained using gene annotation, and the number of reads covering this location was counted. Then, the gene expression level was normalized using the following formula: fragments *per* kilobase million (FPKM) = transcription reads/(transcription length × total mapped reads in the run) × 10^9.

Differentially Expressed Genes Analysis

The normalized FPKM values were used as gene expression levels for the analysis of differentially expressed genes (DEGs) using the Cuffdiff program of the Cufflinks suite (v2.1.1). The differences in gene expression were evaluated using the fold change (≥2.0) and Fisher's exact test (false discovery rate ≤ 0.05).

Functional Annotation of DEGs

For the analysis of Gene Ontology (GO) term enrichment, the DEGs were first annotated with GO terms, and the number of DEGs for each GO term was calculated. Then, the hypergeometric test was used to identify GO terms that were significantly enriched in DEGs when compared to the chicken reference genome. The enrichment was calculated using the following formula: enrichment = $(m/n)/(M/N)$, where N is the total number of genes annotated with a GO term, n is

the number of DEGs in *N*, *M* is the total number of genes annotated with a specific GO term, and *m* is the number of DEGs in *M*. The *p* values were then adjusted by applying the Bonferroni correction, and a *p* value of 0.05 was set as the threshold for adjusted *p* values (false discovery rate). A similar method was used for the analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, except genes were assigned to KEGG pathways instead of being annotated with GO terms.

RT-qPCR Verification of the RNA-Seq Data

RT-qPCR was performed to validate the RNA-Seq results, using the TB Green Premix Ex Taq (Takara, Shiga, Japan) with SYBR Green Dye and the same RNA samples that were used for RNA-Seq. Seven genes were selected for RT-qPCR verification. The primers used for these genes are listed in **Table 1**. The reactions were performed in a total volume of 20 μl according to the manufacturer's recommendations, using an ABI PRISM 7500 sequence detection system (Applied Biosystems, Madrid, Spain) and the following conditions: 5 min at 94°C (1 cycle); 30 s at 94°C, 30 s at annealing temperature (according to the primers listed in **Table 1**), and 30 s at 60°C (35 cycles); and melting curve from 55 to 94°C. Glyceraldehyde 3-phosphate dehydrogenase was selected as the endogenous reference gene, and genes from the L group were set as the criterion. The expression levels were calculated using the $2^{-\Delta\Delta CT}$ method.

Statistical Analysis

All statistical analyses were performed using the SAS 9.3 software (SAS, Cary, NC, USA). Heritability was analyzed using the half-sib correlation method and evaluated using the VARCOMP procedure with the restricted maximum likelihood option. Differences in egg quality among individuals and sire families were compared using the ANOVA procedure. Differences in egg quality between the two consecutive generations were compared using the independent *t* test procedure. The univariate procedure was used to test the normal distribution of the yolk cholesterol content. The general linear model procedure least squares linear model was used to analyze the phenotypic correlation between yolk cholesterol content and egg quality among female individuals and sire families. All data were expressed as mean values ± standard deviation (SD).

RESULTS

Egg Quality and Its Correlation With Yolk Cholesterol Content

Among first-generation female individuals, the egg weight, egg shape index, shell strength, shell thickness, yolk weight, egg white height, Haugh unit, and cholesterol content were 45.36 g, 0.81, 3.07 kg/cm^2, 0.340 mm, 15.57 g, 3.36 mm, 58.70, and 45.86 mmol/L, respectively. Among second-generation female individuals, the egg weight, egg shape index, shell strength, shell thickness, yolk weight, egg white height, Haugh unit, and cholesterol content were 45.16 g, 0.80, 2.97 kg/cm^2, 0.338 mm, 15.57 g, 3.32 mm, 58.42, and 45.25 mmol/L, respectively (**Table 2**). Accordingly, none of the indexes assessed differed significantly between the two generations.

Phenotypic correlation analyses (**Table 3**) suggested that a higher egg weight was associated with an increase in yolk weight, shell strength, shell thickness, and egg white height ($p < 0.05$), and a decrease in the egg shape index ($p < 0.05$). While higher yolk cholesterol was not associated with changes in egg quality among female individuals, higher yolk cholesterol was, however, associated with an increase in yolk weight, egg white height, and yolk color in sire families ($p < 0.05$) of Wenchang chicken.

TABLE 1 | Primers used for RT-qPCR verification of the RNA-Seq data.

No.	Gene symbol	Ensembl accession no.	Primer sequence (5′–3′)	Annealing temperature (°C)
1	CCL19	ENSGALG00000028256	GAAGCTTTAGGGGGAGCCAATCCTCTAAGACCTCTCCGGG	57
2	OSMR	ENSGALG00000003747	TAACTAAAGCAGCGGAGTGCTTTCCCGGGGAGGGTTATCA	55
3	ALOX5	ENSGALG00000005857	CAAACACACGGGAAACCACCCCACCGTCACATCGTAGGAG	57
4	FABP3	ENSGALG00000037050	CCTGGAAGCTGGTGGATACGCCGTGGTCTCATCGAACTCC	59
5	ApoA1	ENSGALG00000007114	GGACCGCATTCGGGATATGGACTTGGCGGAGAACTGGTC	57
6	CYP19A	ENSGALG00000013294	ATGGGGATTGGAAGTGCCTGTCATGAAGAAAGGGCGGACC	57
7	LPL	ENSGALG00000015425	CCCACTGAAACTTTTTCGCCGCTGTCCAGGAACCAGGTAGC	57

TABLE 2 | Egg quality among first- and second-generation female individuals and sire families.

Source	Generation	Egg weight (g)	Egg shape index	Shell strength (kg/cm^2)	Shell thickness (mm)	Yolk weight (g)	Egg white height (mm)	Haugh unit	Cholesterol (mg/egg)
Females	1	45.36 ± 4.44	0.81 ± 0.12	3.07 ± 0.92	0.340 ± 0.032	15.57 ± 1.64	3.36 ± 1.15	58.70 ± 12.33	274.3 ± 36.73
	2	45.16 ± 4.02	0.80 ± 0.08	2.97 ± 0.84	0.338 ± 0.031	15.57 ± 1.57	3.32 ± 0.86	58.42 ± 8.90	265.2 ± 22.88
Sire families	1	44.81 ± 2.89	0.82 ± 0.06	3.07 ± 0.51	0.337 ± 1.72	15.52 ± 0.77	3.36 ± 0.26	58.86 ± 3.08	285.2 ± 128.1
	2	43.88 ± 1.87	0.78 ± 0.02	3.83 ± 0.403	0.373 ± 0.016	13.72 ± 0.61	4.68 ± 0.301	60.20 ± 3.52	282.7 ± 53.5

TABLE 3 | Correlation between the level of cholesterol in egg yolk and egg quality indexes.

Source	Trait	Egg weight	Yolk weight	Egg shape index	Shell thickness	Shell strength	Egg white height	Haugh unit
Females	Cholesterol	0.573	0.978	0.412	0.152	0.432	0.155	0.142
	Egg weight		<0.001	<0.001	<0.001	0.607	0.520	0.101
	Yolk weight			<0.001	<0.001	0.082	0.042	0.238
Sire families	Cholesterol	0.375	<0.001	0.118	<0.001	0.387	<0.001	0.341
	Egg weight		<0.001	<0.001	<0.001	0.006	0.010	0.653
	Yolk weight			<0.001	<0.001	0.747	0.405	0.572

Heritability Evaluation

Among female individuals, the heritability estimates for egg weight, egg shape index, shell strength, shell thickness, yolk weight, and cholesterol content were 0.432, 0.024, 0.030, 0.374, 0.146, and 0.328, respectively (**Table 4**). Among sire families, the heritability estimates for egg weight, egg shape index, shell strength, shell thickness, yolk weight, and cholesterol content were 0.354, 0.070, 0.206, 0.516, 0.176, and 0.530, respectively (**Table 4**). Accordingly, the evaluation of egg weight, shell thickness, and cholesterol content resulted in high heritability estimates for each parameter, while the evaluation of yolk weight and egg shape index resulted in medium and low heritability estimates, respectively. Furthermore, the evaluation of shell strength, shell thickness, and cholesterol content in sire families resulted in higher heritability estimates for each parameter.

RNA-Seq Data and Transcriptome Assembly Results

The sequenced libraries generated an average of 42,290,686 ± 870,109 raw reads *per* library. After filtering using the Q20 standard, the average number of clean reads *per* library was 41,785,132 ± 943,074 with a clean read ratio of 98.80 ± 0.34%. Among the filtered clean reads, an average of 35,157,925 ± 900,332 reads *per* library was mapped to the chicken reference genome with a mapping ratio of 84.14 ± 0.97%. Finally, an average of 28,863,072 ± 981,091 reads *per* library was mapped to genes with a mapping ratio of 85.39 ± 2.25%. The clean reads mapped mostly to gene exons with a ratio of 97.56 ± 0.34%, and according to the sequencing results, an average of 15,613 genes was mapped (**Supplementary Table 1**).

Identification of Candidate Genes Involved in Cholesterol Metabolism

GO Term Analysis of the DEGs

The data from two groups, chickens with the highest and lowest levels of yolk cholesterol, were compared to identify genes with differing reads *per* kilobase *per* million values. Compared to chickens with the lowest level of yolk cholesterol, a total of 375 and 578 genes were down- and upregulated, respectively, in chickens with the highest level of yolk cholesterol (**Figure 1** and **Supplementary Table 2**).

All the DEGs were subjected to GO term and KEGG pathway enrichment analyses. In total, 559 genes were assigned to 2,251 biological processes, 316 cellular components, and 434 molecular functions (**Supplementary Table 3**). Out of these, 42 biological processes, 13 cellular components, and 5 molecular functions were significantly enriched ($p < 0.05$) (**Figure 2**).

Among the various biological processes assigned, positive regulation of response to stimulus (GO:0048584) is the largest category with a total of 749 genes included, and ~13.36% (72 out of 539) of the candidate genes were annotated with this term. Furthermore, two categories of GO terms associated with biological processes were highly represented: GO terms related to cell–cell adhesion (9 GO terms) and the immune response (25 GO terms). Out of these, the GO terms immune system process (GO:0002376) and immune response (GO:0006955) were significantly enriched (**Supplementary Table 3** and **Supplementary Figure 1**). Moreover, the *MYO1G*, *B2M*, *CCL19*, and *CD79B* highly enriched genes were annotated with more than three biological process categories related to the immune response, while the *LCK*, *VAV3*, and *CCLi8* highly enriched genes were annotated with the cell adhesion biological process category (**Supplementary Table 3**).

TABLE 4 | Paternal half-sib family structure and heritability estimates.

Trait	Egg weight (g)	Egg shape index	Shell strength (kg/cm²)	Shell thickness (mm)	Yolk weight (g)	Egg white height (mm)	Haugh unit	Cholesterol (mg/egg)
Sires	24	24	24	24	24	24	24	24
K	7.27	7.27	7.27	7.27	7.27	7.27	7.27	7.27
Progeny	362	362	362	362	362	362	362	362
Heritability								
Females	0.432	0.024	0.030	0.374	0.146	/	/	0.328
Sire families	0.354	0.070	0.206	0.516	0.176	/	/	0.530

$K = (N - \Sigma n_i^2/N)/(S - 1)$, N = total number of progeny, n_i = number of progeny for sire i, and S = number of sires.

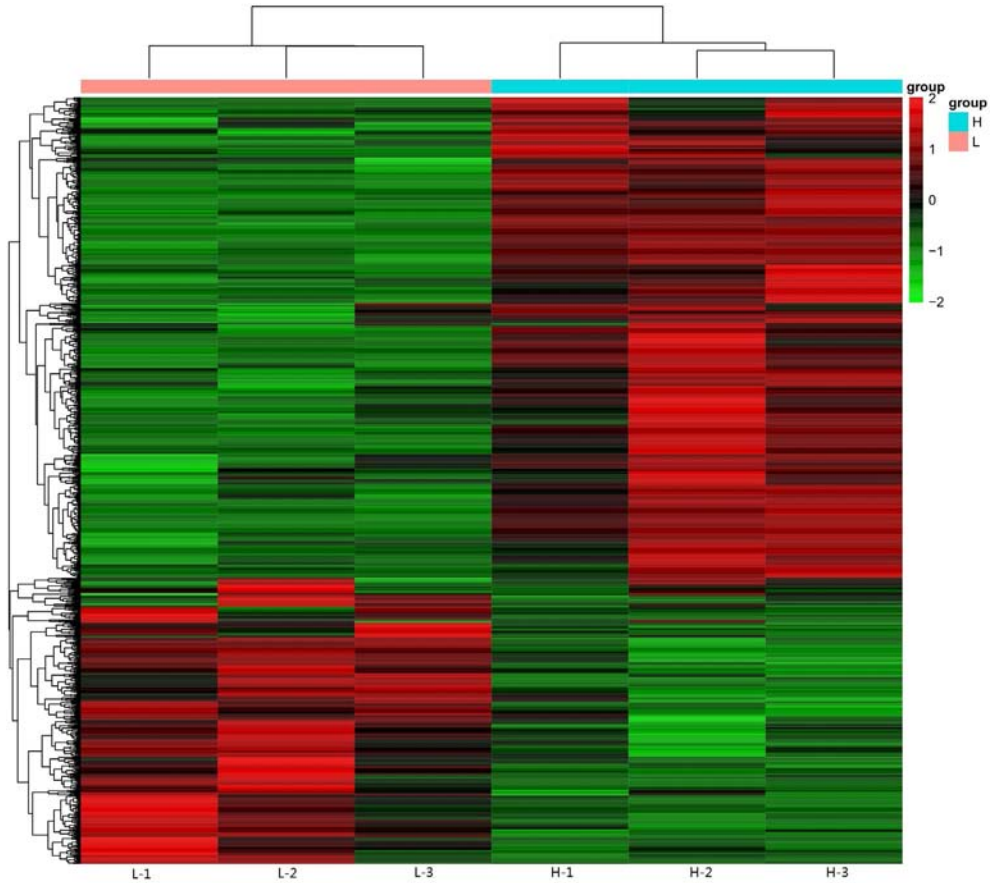


FIGURE 1 | Heatmap analysis of key genes involved in yolk cholesterol deposition. Each row represents a single gene, and each column corresponds to a sequenced sample. The level of expression of each gene is color coded with green and red representing low and high expression levels, respectively.

Regarding cellular component categories, membrane (GO:0016020) and membrane part (GO:0044425) were the two most represented GO terms with 4,104 and 3,114 genes included, respectively. Out of 559 candidate genes, 284 and 228 were assigned to the membrane and membrane part categories, respectively. Furthermore, genes annotated with the GO term condensin complex (GO:0000796) were highly enriched (**Supplementary Figure 2**), and remarkably, all the genes annotated with this GO term were downregulated. Considering the role of transport or secretion through the follicle membrane in cholesterol formation, membrane functions are of particular interest. Among the 13 significantly enriched GO terms for cellular components, 9 are related to the membrane, and 17 enriched genes, including *B2M*, *ALOX5*, *LCP1*, and *LPL*, were annotated with more than 3 membrane-related GO terms.

Lastly, five molecular function categories were enriched (**Supplementary Table 3**), and notably, all the genes annotated with the GO term nonmembrane spanning protein tyrosine kinase activity (GO:0004715) were upregulated. Furthermore, the *CCL4*, *CCL5*, and *CCL19* highly enriched genes were annotated with the signal transport GO term and were all upregulated (**Supplementary Figure 3**).

KEGG Pathway Analysis of the DEGs

In total, 27 KEGG pathways were significantly enriched ($p < 0.05$). They involved 151 genes, 123, and 28 of which were up- and downregulated, respectively. Among the significantly enriched pathways, three were related to signaling interactions and cell transport, and each one of these three pathways involved more than 20 DEGs (**Figure 3** and **Supplementary Table 4**). Furthermore, the highly enriched KEGG pathways were mainly associated with signal transduction, lipid metabolism, and the endocrine system (**Figure 3**). Notably, hematopoietic cell lineage was the most significantly enriched KEGG pathway for the DEGs highly expressed in follicles with the highest level of cholesterol. Moreover, the arachidonic acid metabolism, mineral absorption, PI3K-Akt signaling, ovarian steroidogenesis, and peroxisome proliferator-activated receptors (PPARs) signaling KEGG pathways were involved in the development of follicles with different cholesterol contents. Six genes were involved in ovarian steroidogenesis, among which *CYP2J*, prostaglandin-endoperoxide synthase 2 (*PTGS2*), *ALOX5*, and *ADCY7* were upregulated, while *CYP19A1* and phospholipase A2 group IVF (*PLA2G4F*) were downregulated. Interestingly, *ALOX5* was also annotated with two GO terms (extracellular space and membrane).

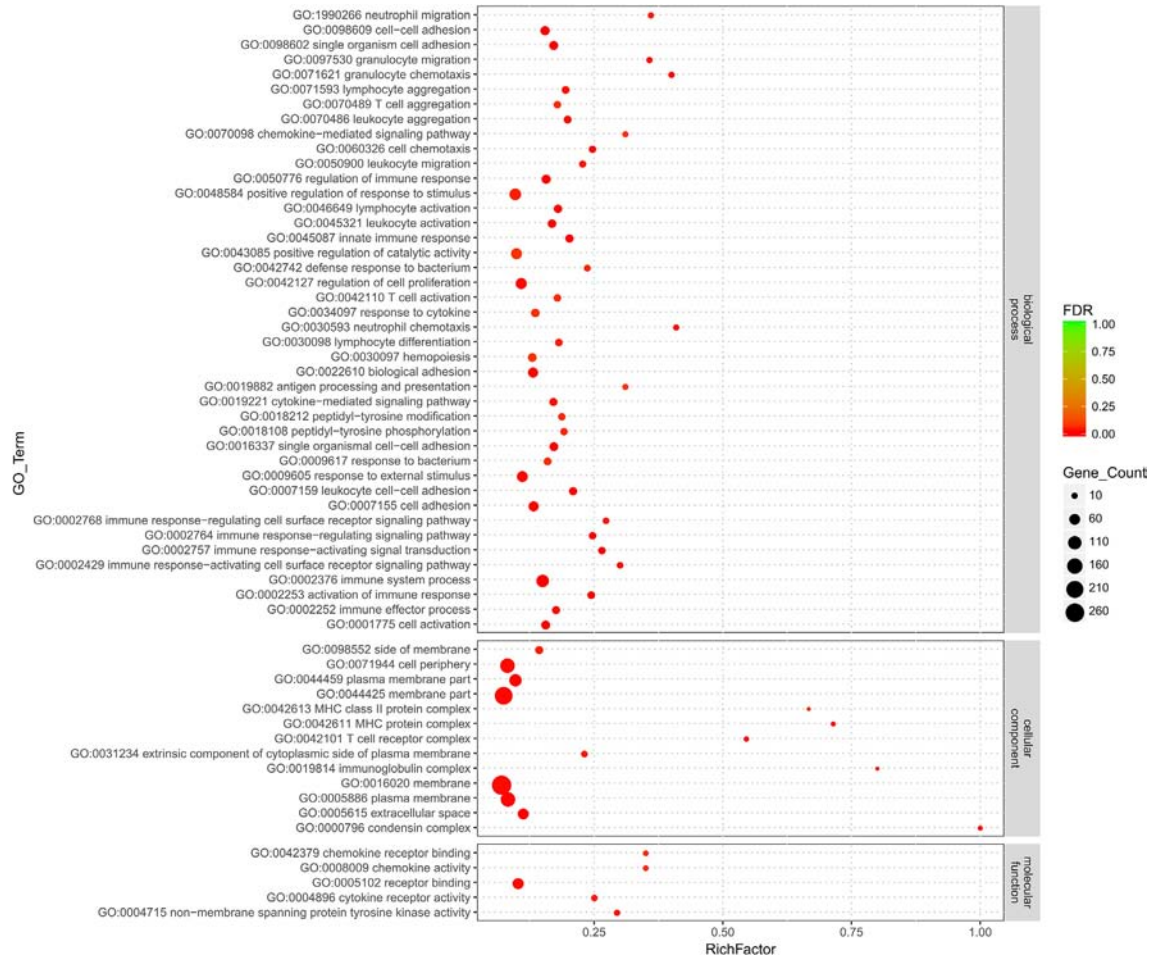


FIGURE 2 | Gene Ontology (GO) term enrichment analysis of candidate genes. The scatter plot presents the results of the GO term enrichment analysis for the candidate genes. The *y*-axis shows the GO terms significantly enriched ($p < 0.05$), and the *x*-axis shows the log 10 *p* values. The size of the bubble corresponding to a specific GO term indicates the number of candidate genes annotated with this term.

Expression of DEGs Involved in the Development of Follicles With Different Cholesterol Contents

We found that many DEGs were involved in the development of follicles with different cholesterol contents, including *B2M*, *ALOX5*, *LCP1*, *LPL*, *FABP3*, *APOA1*, *FLRT2*, *GPRC5B*, *GOLM1*, *GLDN*, and others. Within this list, 23 genes were mapped to the sex chromosome Z, including *LPL*, *CCL19*, *OSMR*, *GOLM1*, and *SYK*.

Next, the highly enriched DEGs were mapped to the chicken protein–protein interaction networks of the STRING database (https://string-db.org). The Cytoscape software was then used to produce a protein–protein interaction plot (**Figure 4**). Lipoprotein lipase (LPL) was significantly downregulated in follicular cells with the highest level of cholesterol and had strong protein–protein interactions, as reflected by high STRING combined scores (the combined score is based on the evidence in the STRING database and reflects the level of confidence of a protein–protein interaction). Meanwhile, *PTGS2* was upregulated and exhibited strong protein–protein interactions (i.e., high STRING combined scores).

We selected seven DEGs (both up- or downregulated in chicken follicular cells with the highest level of cholesterol) and compared the messenger RNA quantification from the transcriptional sequencing results with the expression level assessed by RT-qPCR. Globally, we found a good correlation for the expression trend of the selected genes, as measured by RNA-Seq and RT-qPCR (**Figure 5**). However, the expression of *ALOX5* and *OSMR* exhibited no difference between the H and L groups when measured by RT-qPCR. Furthermore, the detected expression level of *CCL19* was relatively low, while the expression level of *LPL* and *CYP19A* was significantly higher in follicular cells from the L group than in that from the H group.

DISCUSSION

In agreement with a previous report by Baumgartner et al. (2008), this study did not find evidence of a significant association between the yolk cholesterol content and various indexes of egg quality. Accordingly, these observations suggested that the yolk cholesterol content could not be regarded as a standard index for egg quality.

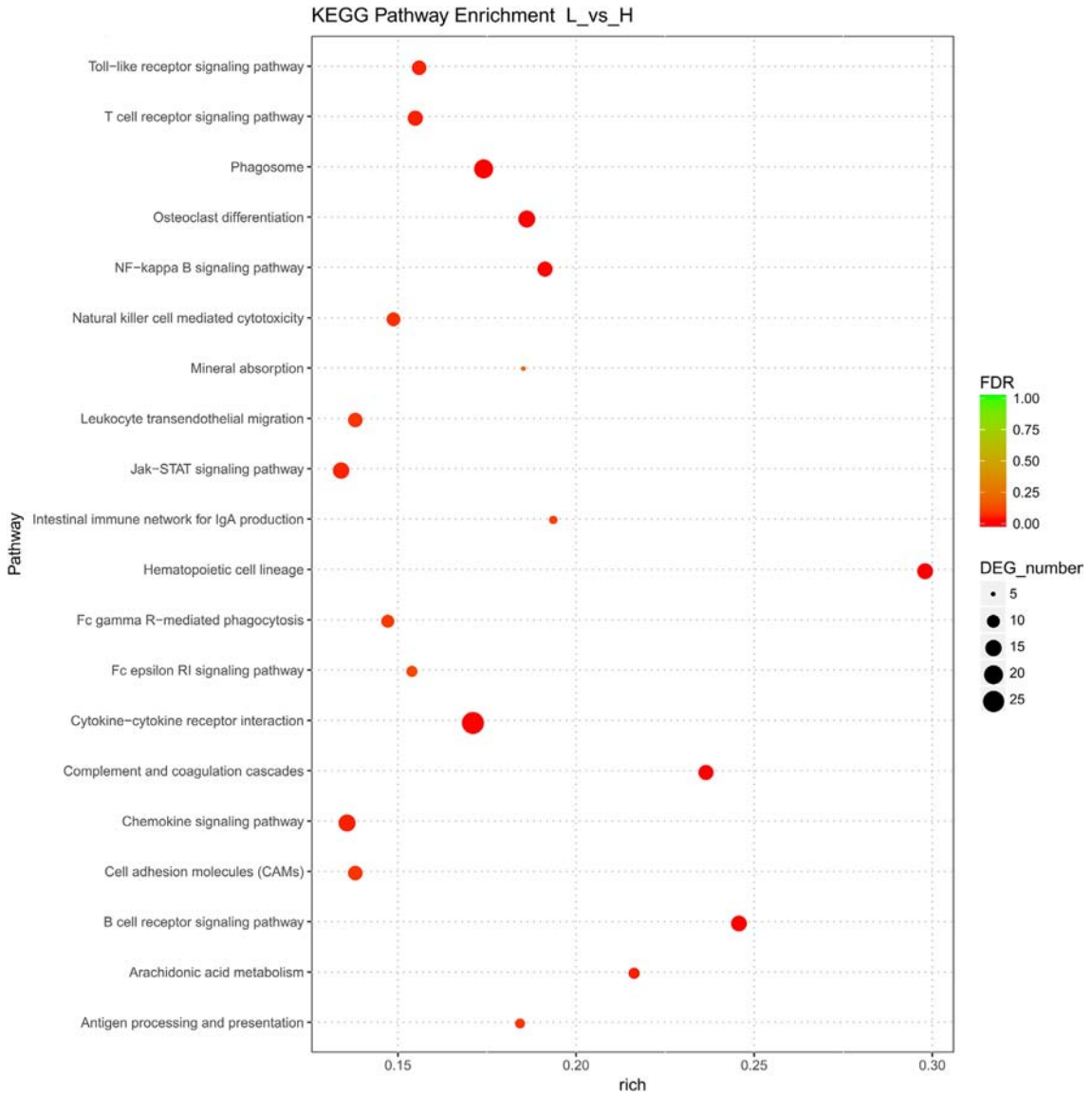


FIGURE 3 | Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of differentially expressed genes. The x-axis shows the enrichment score, the size of a bubble indicates the log value of the number of genes enriched in a pathway, and the color shade represents the *p* value determined using Fisher's exact test.

Furthermore, Ledur et al. (2000) reported that egg quality differed among individuals and increased with age, which suggested that layer performance might be improved by performing selection at an older age. Moreover, the male line is expected to improve egg production at the end of the cycle (Bulut et al., 2013; Goraga et al., 2013). Therefore, it has been proposed that cholesterol synthesis might be affected by the sire family and could be regulated by genes located on chromosome Z (Ledur et al., 2000). Our analyses suggested that the egg weight, shell strength, shell thickness, and egg shape index were correlated with the yolk weight. Indeed, a heavier egg yolk might require more surrounding egg white and shell, which would result in higher egg weight. While the weight of the egg yolk depended on follicular development, the cholesterol content of the egg yolk was positively correlated with the egg weight, which suggested that cholesterol and egg yolk were the most important factors affecting egg weight (Baumgartner et al., 2008). In general, a relatively high cholesterol content has been associated with good health conditions in birds, whereas higher nutrient content in the egg yolk has been associated with a higher egg weight (Zhang, 2016).

In this study, the heritability estimate for the egg weight in Wenchang chickens was 0.432 in females and 0.354 in sire families. Overall, these estimates are in agreement with a previous study by Rath et al. (2015), which reported a heritability estimate of 0.443 for the egg weight in white leghorns chickens. Furthermore, in this study, shell thickness was positively correlated with shell weight, and the estimated heritability of shell strength (0.030 in females and 0.206 in sire families) was consistent with previous reports (Rath et al., 2015; Alwell et al., 2018). For the moderate heritability

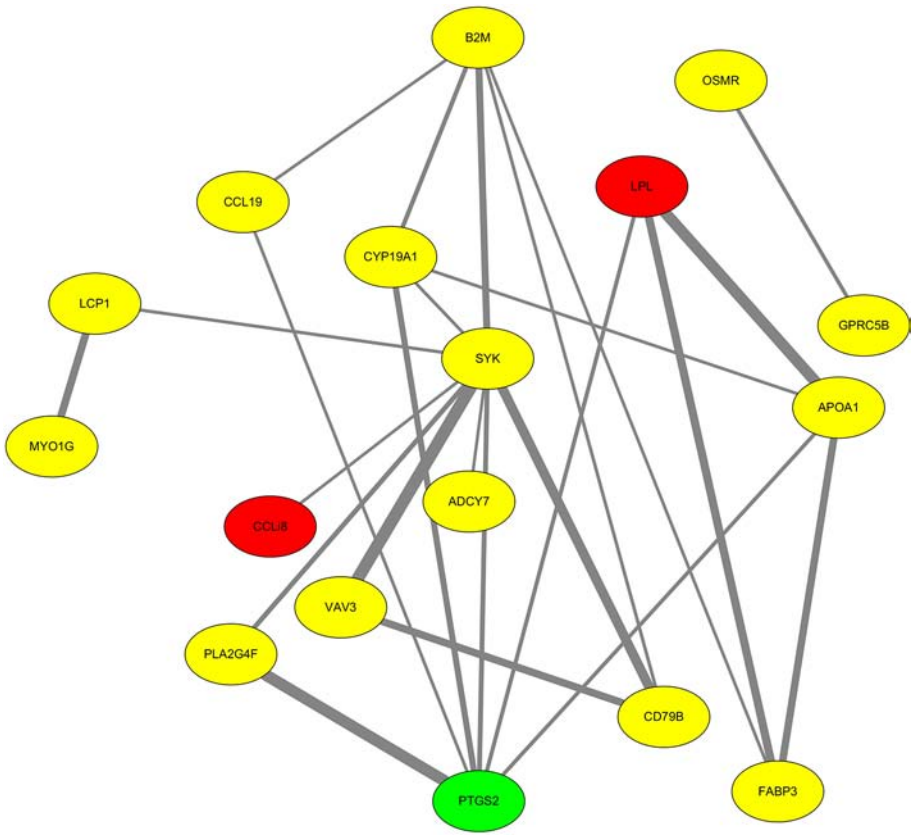


FIGURE 4 | Protein–protein interaction network analysis of selected differentially expressed genes. Proteins highlighted in red and green were significantly down- and upregulated, respectively, while proteins highlighted in yellow showed no significant difference. In the network, each line represents the strength of the relationship between two proteins. Strong interactions are indicated by high STRING combined scores and wide lines, while weak interactions are indicated by low STRING combined scores and narrow lines.

of shell strength in sire family, it might be more appropriate for sire selection to achieve a quick progress in breeding. In contrast, the egg shape index and yolk weight had relatively low heritability estimates, which might be due to the high phenotypic variance, and further suggested that these two traits could not be selected using phenotypic values. Lastly, the heritability estimate for yolk cholesterol content was higher in sire families than in females, which further indicated that yolk cholesterol content was controlled by genes located on the chromosome Z and could be selected through the male line (Ledur et al., 2000).

Ovarian follicle development requires markedly increased DNA and protein synthesis in the granulosa cells of the follicle membrane (Seol et al., 2006; Bonnet et al., 2011). During the rapid growth of chicken follicles, DNA and protein synthesis is stimulated and regulated by a variety of steroid hormones (Diaz, 2011) and the expression of genes involved in this progress. For example, the biological function of the phospholipase A2 (PLA2) subfamily of enzymes is to catalyze the hydrolysis of the sn-2 position of membrane glycerophospholipids, which leads to the production of free fatty acids and lysophospholipids (Duncan et al., 2008). Furthermore, several reports have involved PLA2 in the induction of cell apoptosis. In chickens with the highest yolk cholesterol content, the downregulation of *PLA2G4* in the ovarian steroidogenesis pathway suggested that increased phospholipids synthesis was required for cholesterol deposition (Diouf et al., 2006; Aljakna et al., 2012). Moreover, PTGS2 has been reported to be induced or upregulated by the luteinizing hormone surge during ovulation in rodent and fish (Yerushalmi et al., 2014; Tang et al., 2017). Therefore, the upregulation of *PTGS2* in follicles might also suggest that ovulation occurs more frequently in chickens producing eggs with a higher cholesterol content. Indeed, the increased level of PTGS2, together with the action of arachidonate-5-lipoxygenase (ALOX5), would further promote the release of arachidonic acid (Kurusu et al., 2009), and the subsequent conversion of arachidonic acid by downstream metabolic enzymes of the CYP2J subfamily could impact the ovulatory mechanisms (Newman et al., 2004).

The expression of LPL in the ovarian follicles of domestic chicken was first identified by Benson et al. (1975). LPL is an essential enzyme of VLDL metabolism and exhibits high levels of expression in rapidly growing ovarian follicles, which provides follicular tissues with the enzyme required to hydrolyze VLDL into fatty acids and monoglycerides (Gupta et al., 2017). In the present study, LPL was expressed at a relatively low level in the ovarian follicles with the highest cholesterol content. Therefore, we would like to propose that low levels of LPL play a role in the retention

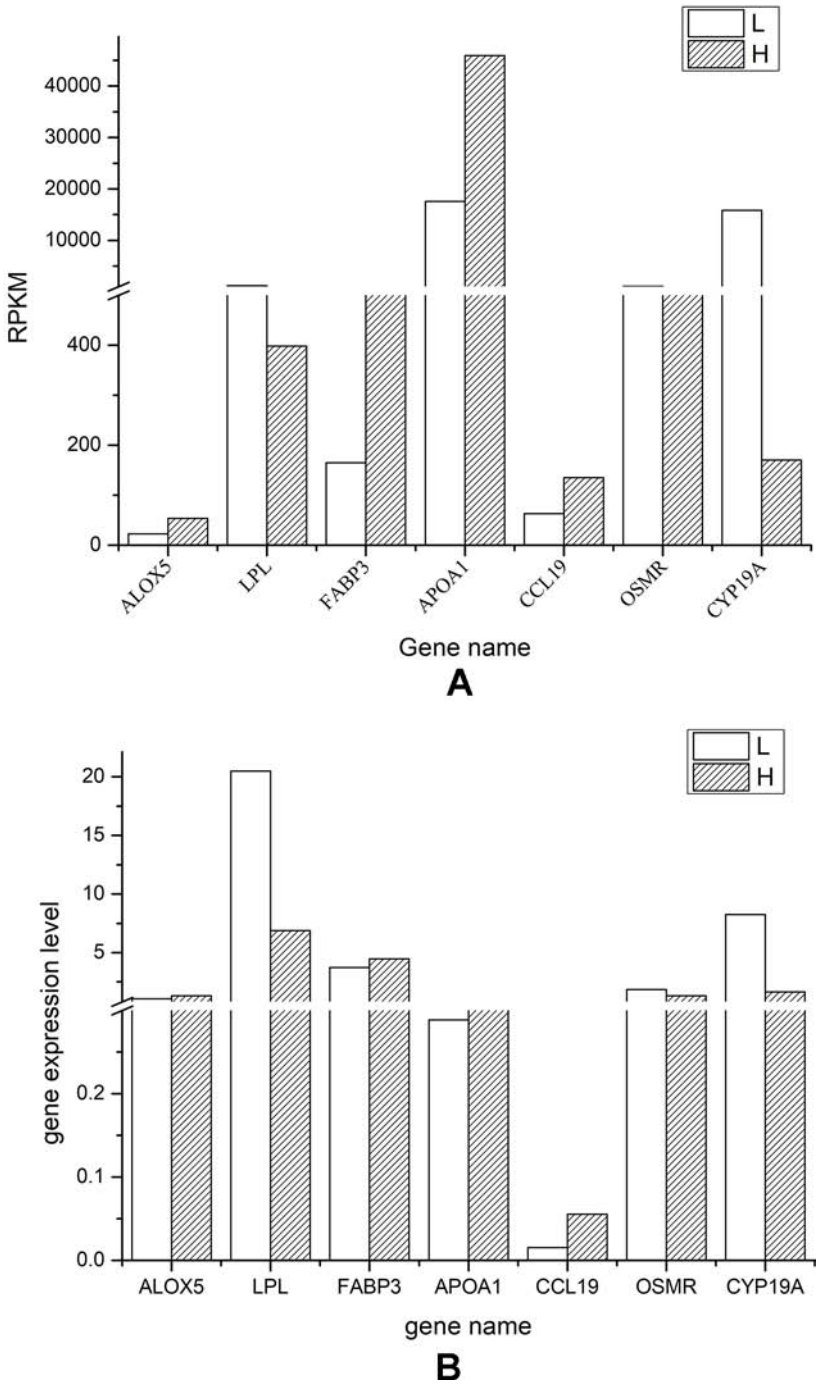


FIGURE 5 | Validation of the RNA-Seq results *via* real-time quantitative PCR RT-qPCR analyses. **(A)** Diagram showing the reads *per* kilobase *per* million value of each gene in both low (L) and high (H) cholesterol content groups. **(B)** Diagram showing the expression level quantified by RT-qPCR of the indicated genes in both L and H groups.

of high VLDL levels, which in turn leads to an increase in the amount of VLDL-cholesterol and triglyceride-rich lipoproteins in ovarian follicles. Furthermore, VLDL has been demonstrated to be a source of neutral lipids in the oocytes of anguillid eels and cutthroat trouts (Damsteegt et al., 2015; Lubzens et al., 2017). Moreover, the downregulation of *LPL* has been involved in the PPAR signaling pathway. As part of the PPAR signaling pathway, ApoA1 and FABP3 play roles in lipid metabolism (Wang et al., 2017b), while PEPCK plays a role in gluconeogenesis (Glorian et al., 2001). ApoA1, FABP3, and PEPCK are also all upregulated in response to retinoid X receptor alpha. Unlike mammalians where females have XX and males XY sex chromosomes, birds

have the ZW system where females have ZW and males ZZ sex chromosomes. In male chickens, it has been shown that the two copies of chromosome Z are not affected by global dosage compensation mechanisms, and therefore, genes located on chromosome Z usually exhibit higher levels of expression in males than in females (Toups et al., 2011). The *LPL* gene is assigned to chromosome Z and usually exhibits a low level of expression in birds (Han, 2005), which might explain the negative correlation with the yolk cholesterol content in sire families.

In mammals, pregnancy will improve the innate and adaptive immunity during gestation to increase pregnancy outcomes (Kraus et al., 2012). Similar to mammals' pregnancy, follicles formation and ovulation in chickens may also need improved immunity to guarantee a higher egg quality. During the rapid growth phase of ovarian follicles, the components of the follicle matrix expand rapidly, which acts as intrinsic mechanical stress during the accumulation of yolk precursors (Kraus et al., 2012; Richards et al., 2008). We speculated that, in the follicles with the highest cholesterol content, this phenomenon was responsible for the increased expression of genes related to the immune response and signaling pathways, including hematopoietic cell lineage, toll-like receptor signaling pathway, and others.

Energy and substrate sources are also required for ovarian folliculogenesis (Seol et al., 2006). Interestingly, genes related to the arachidonic acid metabolism, which contributes to energy intake, were significantly enriched in ovarian follicles with a high cholesterol content (Lee et al., 2005). Our data suggested that VLDL absorption as a yolk precursor in ovarian follicles with the highest cholesterol content was mediated through the downregulation of *LPL* expression. This contrast with the situation in mammals, where phospholipase A2 group IVA (*PLA2G4A*) expression is upregulated in granulosa cells at ovulation (Diouf et al., 2006), and the yolk exhibits a higher content of arachidonic acid through the down- and upregulation of *PLA2G4F* and *PTGS2*, respectively. Furthermore, in cows, the upregulation of *PLA2G4A* has been associated with a down- and upregulation of *CYP19A1* and *PTGS2*, respectively (Sirois, 1994). These differences might indicate that a high cholesterol content requires arachidonic acid degradation and *PLA2* downregulation to maintain high levels of phospholipids while keeping the same expression trend for *CYP19A1* and *PTGS2*.

CONCLUSIONS

The yolk cholesterol content was most affected by the sire family with a heritability estimate of 0.530. Furthermore, the ovarian steroidogenesis pathway appeared to affect the yolk cholesterol content, with the downregulation of the *LPL* gene located on chromosome Z playing key roles. In contrast to mammals, a high yolk cholesterol content appeared to require the downregulation of *PLA2G4A* in chickens, which might also affect ovulation. Nevertheless, further studies with *LPL* overexpression or knockdown are required to confirm its role in the functional regulation of the yolk cholesterol content in birds.

DATA AVAILABILITY STATEMENT

The data used in this manuscript can be found according to the link below: https://www.ncbi.nlm.nih.gov//bioproject/PRJNA532290.

ETHICS STATEMENT

All experimental procedures were performed following guidelines developed by the China Council on Animal Care and Protocols and were approved by the Animal Care and Use Committee of Anhui Agricultural University, China (permission No. SYDW-P2017062801).

AUTHOR CONTRIBUTIONS

XC designed the study, analyzed and interpreted the data, and wrote the paper. WZ conducted egg quality measurement and follicle membrane collection. YD conducted qPCR experiments. XL extracted RNA from follicle membrane. ZG designed the study.

FUNDING

Support for this project was provided in part by the Major Scientific and Technological Special Project in Anhui Province (18030701174), the Open Fund of Anhui Province Key Laboratory of Local Livestock and Poultry, Genetical Resource Conservation and Breeding (AKLGRCB2017001), and the Key project of natural fund of Anhui Provincial Education Department (KJ2018A951).

ACKNOWLEDGMENTS

We thank Ms. Boni G. Funmilayo, from Anhui Province Key Laboratory of Local Livestock and Poultry Genetic Resource Conservation and Bio-breeding, for correcting the grammar of the manuscript. We also thank Anhui Huadong Mountain Fresh Agricultural Development Co., Ltd for providing all the experimental materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00902/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Directed acyclic graph (DAG)display of GO highly enriched biological process results with candidate targeted genes. The enrichment of GO terms is color coded from low (light yellow) to high (red).

SUPPLEMENTARY FIGURE 2 | Directed acyclic graph (DAG) display of GO highly enriched cellular component results with candidate targeted genes. The enrichment of GO terms is color coded from low (light yellow) to high (red).

SUPPLEMENTARY FIGURE 3 | Directed acyclic graph (DAG) display of GO highly enriched molecular function results with candidate targeted genes. The enrichment of GO terms is color coded from low (light yellow) to high (red).

REFERENCES

Aljakna, A., Choi, S., Savage, H., Blair, R. H., Gu, T., Svenson, K. L., et al. (2012). Pla2g12b and Hpn are genes identified by mouse ENU mutagenesis that affect HDL cholesterol. *PLoS One* 7, e43139. doi: 10.1371/journal.pone.0043139

Alwell, J. S., Abdur-Rahman, A., and Chukwujindu, N. S. (2018). Heritability estimates of egg weight and egg shell weight in Ikenne, Nigeria. *Intl. J. Sci. World* 6, 38–42. doi: 10.14419/ijsw.v6i1.8677

Andersen, C. J., Blesso, C. N., Lee, J., Barona, J., Shah, D., Thomas, M. J., et al. (2013). Egg consumption modulates HDL lipid composition and increases the cholesterol-accepting capacity of serum in metabokic syndrome. *Lipids* 48, 557–567. doi: 10.1007/s11745-013-3780-8

Baumgartner, J., Koncekova, Z., Benkova, J., Peskovicova, D., Simenovova, J., and Csuka, J. (2008). Changes in egg quality traits associated with long-term selection for lower yolk cholesterol content in Japanese quail. *Czech J. Anim. Sci.* 53, 119–127. doi: 10.17221/2715-CJAS

Benson, J. D., Bensadoun, A., and Cohen, D. (1975). Lipoprotein lipase of ovarian follicles in the domestic chicken (*Gallus* domesticus). *Proc. Soc. Exp. Biol. Med.* 148, 347–350. doi: 10.3181/00379727-148-38537

Bonnet, A., Bevilacqua, C., Benne, F., Bodin, L., Cotinot, C., Liaubet, L., et al. (2011). Transcriptome profiling of sheep granulose cells and oocytes during early follicular development obtained by Laser Capture Microdissection. *BMC Genomics* 12, 417. doi: 10.1186/1471-2164-12-417

Bulut, Z., Kurar, E., Ozsensoy, Y., Nizamlioglu, M., Garip, M., Yilmaz, A., et al. (2013). Determination of chromosomal regions affecting body weight and egg production in Denizli × White Leghorn F2 populations. *Eurasian J. Vet. Sci.* 29, 30–38.

Damsteegt, E. L., Falahatimarvast, A., Mccormick, S. P., and Lokman, P. M. (2015). Triacylglyceride physiology in the short-finned eel, Angulla australis—changes throughout early oogenesis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 308, R935–R944. doi: 10.1152/ajpregu.00436.2014

Diaz, F. J. (2011). Early avian follicular development is characterized by changes in transcripts involved in steroidogenesis, paracrine signaling and transcription. *Mol. Reprod. Dev.* 78, 212–223. doi: 10.1002/mrd.21288

Dikmen, B. Y., and Sahan, U. (2007). Correlations between breeder age, cholesterol content, blood cholesterol level and hatchability of broiler breeders. *Br. Poult. Sci.* 48, 98–103. doi: 10.1080/00071660601161412

Ding, X., Yu, Y., Su, Z., and Zhang, K. (2017). Effects of essential oils on performance, egg quality, nutrient digestibility and yolk fatty acid profile in laying hens. *Anim. Nutr.* 3, 127–131. doi: 10.1016/j.aninu.2017.03.005

Diouf, M. N., Sayasith, K., Lefebvre, R., Silversides, D. W., Sirois, J., and Lussier, J. G. (2006). Expression of phospholipase A2 group IVA (PLA2G4A) is upregulated by human chorionic gonadotropin in bovine granulose cells of ovulatory follicles. *Biol. Reprod.* 74, 1096–1103. doi: 10.1095/biolreprod.105.048579

Duncan, R. E., Sarkadi-Nagy, E., Jaworski, K., Ahmadian, M., and Sul, H. S. (2008). Identification and functional characterization of adipose-specific phospholipase A2 (AdPLA). *J. Biol. Chem.* 283, 25428–25436. doi: 10.1074/jbc.M804146200

Glorian, M., Duplus, E., Beale, E. G., Scott, D. K., Granner, D. K., and Forest, C. (2001). A single element in the phosphoenolpyruvate carboxykinase gene mediates thiazolidinedione action specifically in adipocytes. *Biochimie* 83, 933–943. doi: 10.1016/S0300-9084(01)01343-8

Gohin, M., Bobe, J., and Chesnel, F. (2010). Comparative transcriptomic analysis of follicle-enclosed oocyte maturational and developmental competence acquisition in two non-mammalian vertebrates. *BMC Genomics* 11, 18. doi: 10.1186/1471-2164-11-18

Goraga, Z. S., Nassar, M. K., and Brockmann, G. A. (2013). Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: I. *Egg production traits. Anim. Genet.* 44, 62–68. doi: 10.1111/j.1365-2052.2012.02365.x

Griffin, H. D. (1992). Manipulation of egg yolk cholesterol: a physiologist's view. *Worlds Poult. Sci. J.* 48, 101–112. doi: 10.1079/WPS19920010

Gupta, A., Tiwari, M., Prasad, S., and Chaube, S. K. (2017). Role of cyclic nucleotide phosphodiesterases during meiotic resumption from diplotene arrest in mammalian oocytes. *J. Cell. Biochem.* 118, 446–452. doi: 10.1002/jcb.25748

Han, D. (2005). *Molecular approaches to understanding variation in reproductive phenotype of female zebra finches (Taeniopygia guttata)*. Beijing (BJ): Peking University.

Janjira, S. (2017). Preliminary study: egg production performance, egg quality and blood plasma cholesterol concentration in laying hens fed dietary dried fermented ginger and/or fermented corncob powder. *Food Sci. Nutr.* 3, 1–5. doi: 10.24966/FSN-1076/100016

Klkin, R. G., Yan, Z., Buhman, K. K., Story, J. A., Turek, J. J., Anderson, M., et al. (1997). Reduction of egg yolk cholesterol content through inhibition of hepatic cholesterol biosynthesis and alteration of plasma VLDL composition in laying hens: comparative effects of atorvastatin, lovastatin, and simvastatin. *Atherosclerosis* 134, 123. doi: 10.1016/S0021-9150(97)88669-8

Kraus, T. A., Engel, S. M., Sperling, R. S., Kellerman, L., Lo, Y., Wallenstein, S., et al. (2012). Characterizing the pregnancy immune phenotype: results of the viral immunity and pregnancy (VIP) study. *J. Clin. Immunol.* 32, 300–311. doi: 10.1007/s10875-011-9627-2

Kurusu, S., Jinno, M., Ehara, H., Yonezawa, T., and Kawaminami, M. (2009). Inhibition of ovulation by a lipoxygenase inhibitor involves reduced cyclooxygenase-2 expression and prostaglandin E2 production in gonadotropin-primed immature rats. *Reproduction* 137, 59–66. doi: 10.1530/REP-08-0257

Ledur, M. C., Fairfull, R. W., Mcmillan, I., and Asseltine, L. (2000). Genetic effects of aging on egg production traits in the first laying cycle of white Leghorn strains and strain crosses. *Poultry Sci.* 79, 296–304. doi: 10.1093/ps/79.3.296

Lee, Y., Kim, H., Kim, M., and Chun, S. (2005). Control mechanisms of ovulation by pituitary adenylate cyclase-activating polypeptide. *Korean J. Fertil. Steril.* 32, 101–112.

Lubzens, E., Bobe, J., Young, G., and Sullivan, C. V. (2017). Maternal investment in fish oocytes and eggs: The molecular cargo and its contributions to fertility and early development. *Aquaculture* 472, 107–143. doi: 10.1016/j.aquaculture.2016.10.029

Newman, J. W., Stok, J. E., Vidal, J. D., Corbin, C. J., Huang, Q., Hammock, B. D., et al. (2004). Cytochrome p450-dependent lipid metabolism in preovulatory follicles. *Endocrinology* 145, 5097–5105. doi: 10.1210/en.2004-0710

Omole, J. O., and Ighodaro, O. M. (2013). Comparative studies of the effects of egg yolk, oats, apple, and wheat bran on serum lipid profile of Wistar rats. *ISRN Nutr.* 2013, 730479. doi: 10.5402/2013/730479

Panda, A. K., Reddy, M. R., Rama Rao, S. V., and Praharaj, N. K. (2003). Production performance, serum/yolk cholesterol and immune competence of White Leghorn layers as influenced by dietary supplementation with probiotic. *Trop. Anim. Health Prod.* 35, 85–94. doi: 10.1023/A:1022036023325

Rath, K. P., Prasanna, K. M., Bandi, K. M., and Nrusingha, C. B. (2015). Evaluation of different egg quality traits and interpretation of their mode of inheritance in white leghorn. *Vet. World* 8, 449–452. doi: 10.14202/vetworld.2015.449-452

Richards, J. S., Liu, Z., and Shimada, M. (2008). Immune-like mechanisms in ovulation. *Trends Endocrinol. Metabol.* 19, 191–196. doi: 10.1016/j.tem.2008.03.001

Seol, H. S., Sato, K., Murakami, H., Toyomizu, M., and Akiba, Y. (2006). Changes in gene expression involved in energy utilization during chicken follicle development. *Anim. Reprod. Sci.* 95, 283–294. doi: 10.1016/j.anireprosci.2005.09.016

Sirois, J. (1994). Induction of prostaglandin endoperoxide synthase-2 by human chorionic gonadotropin in bovine preovulatory follicles *in vivo. Endocrinology* 135, 841–848. doi: 10.1210/endo.135.3.8070377

Sreenivas, D., Prakash, G. M., Mahender, M., and Chatterjee, R. N. (2013). Genetic analysis of egg quality traits in White Leghorn chicken. *Vet. World* 6, 263–266. doi: 10.5455/vetworld.2013.263-266

Su, Y., Sugiura, K., Wigglesworth, K., O'Brien, M. J., Affourtit, J. P., Pangas, S. A., et al. (2008). Oocyte regulation of metabolic cooperativity between mouse cumulus cells and oocytes: BMP15 and GDF9 control cholesterol biosynthesis in cumulus cells. *Development* 135, 111–121. doi: 10.1242/dev.009068

Tang, H., Liu, Y., Li, J., Li, G., Chen, Y., Yin, Y., et al. (2017). LH signaling induced ptgs2a expression is required for ovulation in zebrafish. *Mol. Cell. Endocrinol.* 447, 125–133. doi: 10.1016/j.mce.2017.02.042

Toups, M. A., Pease, J. B., and Hahn, M. W. (2011). No excess gene movement is detected off the avian or lepidopteran Z chromosome. *Genome Biol. Evol.* 3, 1381–1390. doi: 10.1093/gbe/evr109

Wang, Z., Meng, G., Bai, Y., Liu, R., Du, Y., and Su, L. (2017a). Comparative transcriptome analysis provides clues to molecular mechanisms underlying

blue-green eggshell color in the Jinding duck (*Anas platyrhynchos*). *BMC Genomics* 18, 725. doi: 10.1186/s12864-017-4135-2

Wang, Z., Shang, P., Li, Q., Wang, L., Chamba, Y., Zhang, B., et al. (2017b). iTRAQ-based proteomic analysis reveals key proteins affecting muscle growth and lipid deposition in pigs. *Sci. Rep.* 7, 46717. doi: 10.1038/srep46717

Yadgary, L., Cahaner, A., Kedar, O., and Uni, Z. (2010). Yolk sac nutrient composition and fat uptake in late-term embryos in eggs from young and old broiler breeder hens. *Poultry Sci.* 89, 2441–2452. doi: 10.3382/ps.2010-00681

Yair, R., and Uni, Z. (2011). Content and uptake of minerals in the yolk of broiler embryos during incubation and effect of nutrient enrichment. *Poultry Sci.* 90, 1523–1531. doi: 10.3382/ps.2010-01283

Yang, P., Tian, Y., Sun, G., Jiang, R., Han, R., and Kang, X. (2013). Deposition rule of yolk cholesterol in two different breeds of laying hens. *Genet. Mol. Res.* 12, 5786–5792. doi: 10.4238/2013.November.22.5

Yerushalmi, G. M., Salmondivon, M., Yung, Y., Maman, E., Kedem, A., Ophir, L., et al. (2014). Characterization of the human cumulus cell transcriptome during final follicular maturation and ovulation. *Mol. Hum. Reprod.* 20, 719–735. doi: 10.1093/molehr/gau031

Zhang, Y. (2016). *Studies on chicken hatchability and its relation with egg yolk metabolites*. Niedersachsen (NI): Georg-August-University Göttingen.

Zhou, L., Shi, Y., Guo, R., Liang, M., Zhu, X., and Wang, C. (2014). Digital gene expression profiling analysis of the cholesterol-lowering effects of alfalfa saponin extract on laying hens. *PLoS One* 9, e98578. doi: 10.1371/journal.pone.0098578

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Zhu, Du, Liu and Geng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 October 2019
doi: 10.3389/fgene.2019.00993

Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response

*Krishnamoorthy Srikanth[1], Himansu Kumar[1], Woncheoul Park[1], Mijeong Byun[1], Dajeong Lim[1], Steve Kemp[2], Marinus F. W. te Pas[3], Jun-Mo Kim[4] and Jong-Eun Park[1]**

[1] Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Wanju, South Korea, [2] Animal Biosciences, International Livestock Research Institute (ILRI), Nairobi, Kenya, [3] Wageningen UR Livestock Research, Animal Breeding and Genomics, Wageningen, Netherlands, [4] Department of Animal Science and Technology, Chung-Ang University, Anseong, South Korea

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Jian Xu, Chinese Academy of Fishery Sciences (CAFS), China
Luyang Sun, Baylor College of Medicine, United States

***Correspondence:**
Jong-Eun Park
jepark0105@korea.kr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *17 May 2019*
Accepted: *18 September 2019*
Published: *11 October 2019*

Citation:
Srikanth K, Kumar H, Park W, Byun M, Lim D, Kemp S, te Pas MFW, Kim J-M and Park J-E (2019) Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response. Front. Genet. 10:993. doi: 10.3389/fgene.2019.00993

Heat stress (HS) negatively affects chicken performance. Agricultural expansion will happen in regions that experience high ambient temperatures, where fast-growing commercial chickens are vulnerable. Indigenous chickens of such regions, due to generations of exposure to environmental challenges, might have higher thermal tolerance. In this study, two indigenous chicken ecotypes, from the hot and humid Mombasa (lowland) and the colder Naivasha (highland) regions, were used to investigate the effects of acute (5 h, 35°C) and chronic (3 days of 35°C for 8 h/day) HS on the cardiac and skeletal muscle, through RNA sequencing. The rectal temperature gain and the number of differentially expressed genes (DEGs) [False Discovery Rate (FDR) < 0.05] were two times higher in the acute stage than in the chronic stage in both ecotypes, suggesting that cyclic exposure to HS can lead to adaptation. A tissue- and stage-specific difference in response to HS was observed, with peroxisome proliferator-activated-receptor (PPAR) signaling and mitogen-activate protein kinase (MAPK) signaling pathways, enriched in heart and skeletal muscle, respectively, and the p53 pathway enriched only in the acute stage in both tissues. The acute and chronic stage DEGs were integrated by a region-specific gene coexpression network (GCN), and genes with the highest number of connections (hub genes) were identified. The hub genes in the lowland network were *CCNB2*, *Crb2*, *CHST9*, *SESN1*, and *NR4A3*, while *COMMD4*, *TTC32*, *H1F0*, *ACYP1*, and *RPS28* were the hub genes in the highland network. Pathway analysis of genes in the GCN showed that p53 and PPAR signaling pathways were enriched in both low and highland networks, while MAPK signaling and protein processing in endoplasmic reticulum were enriched only in the gene network of highland chickens. This shows that to dissipate the accumulated heat, to reduce heat induced apoptosis, and to promote DNA damage repair, the ecotypes activated or suppressed different genes, indicating the differences in thermal tolerance and HS response mechanisms between the

ecotypes. This study provides information on the HS response of chickens, adapted to two different agro climatic environments, extending our understanding of the mechanisms of HS response and the effect of adaptation in counteracting HS.

Keywords: heat stress, hub genes, PPAR signaling, MAPK signaling, p53 signaling, RNA-Seq

INTRODUCTION

Chicken is a cheap source of high-quality protein and provides significant food and income security for rural communities (Mekonnen et al., 2010). Specialized trait selection, under controlled environments, has made commercial broiler chickens sensitive to environmental extremes (Mcmichael et al., 2007; Ciscar et al., 2011; Kantanen et al., 2015). This creates significant hindrance for the expansion of poultry industry into regions that experience environmental conditions such as heat stress (HS) (Canario et al., 2013; Lawrence and Wall, 2014; Rothschild and Plastow, 2014). Being homeothermic, chickens are able to maintain a constant body temperature across a wide range of temperature (Deeb and Cahaner, 1999); however, increasing ambient temperature due to global warming and climate change will have a major impact on the animal's physiology and performance, resulting in significant economic losses to livestock industries (Renaudeau et al., 2012; Wang et al., 2017). HS is classified as the state at which the ambient temperature exceeds the tolerable range, making it difficult for the birds to maintain its homeostatic body temperature (Lara and Rostagno, 2013). It leads to reduction in meat quality, low growth rate, reductions in body weight, reduced egg weight and shell thickness, and also high mortality in commercial layers and boilers (Muiruri and Harrison, 1991; Wolfenson et al., 2001). They also cause significant immunosuppression due to reduced humoral immunity, rendering the birds susceptible to diseases (Padgett and Glaser, 2003; Wang et al., 2017; Monson et al., 2018). The birds' response to HS depends on its genetics (Felver-Gant et al., 2012). While commercial fast-growing broilers are particularly more sensitive to HS (Yunis and Cahaner, 1999), indigenous chicken (IC) breeds that are native to tropical areas are documented to have higher HS tolerance relative to other breeds (Soleimani and Zulkifli, 2010), suggesting that genetic resistance to HS can be acquired as a consequence of adaptation and can be inherited (Lu et al., 2007).

Future agricultural expansion, to support increasing global population, will mainly happen in regions with climatic conditions that are less suitable for commercial livestock, which lack the genetic potential to adapt to environmental extremes (Lara and Rostagno, 2013; Porto-Neto et al., 2014; Rothschild and Plastow, 2014). In the villages of developing countries like Kenya, IC production provides not only food security but also income security due to their low production cost and their ability to survive on scavenging and their resilience to environmental parasite challenge (Magothe et al., 2012). It was reported that 70% out of a total of 31.8 million domesticated chickens in Kenya were IC (Moraa et al., 2015). Kenya has seven different agroecological zones, including, arid, semi-arid, tropical, and temperate regions (Silvestri et al., 2012). Twelve ecotypes of chickens are found across these agroecological zones (Kingori et al., 2010; Moraa et al., 2015). The chickens show high genetic diversity and are well adapted to their local environment (Nyaga, 2007). The chickens are raised extensively under free range systems (Sonaiya, 1990), which exposes them to the negative influence of extreme weather changes. Native chicken ecotypes that have survived extreme environmental conditions over multiple generations would have developed tolerance at the genomic level (Chen et al., 2014; Lawrence and Wall, 2014; Porto-Neto et al., 2014; Fleming et al., 2017). Therefore, to mitigate impacts of HS through genetic approaches, it is prudent to examine chickens that have evolved in such environments (Fleming et al., 2017). Kenyan IC presents an opportunity to understand the genetic response to HS of hot temperature–adapted chickens. Previous studies on HS adapted and nonadapted chickens revealed the biological mechanisms regulated by HS and identified differential immune response between lowland- and highland-adapted chickens exposed to tropical conditions (Park et al., 2019; Te Pas et al., 2019). In this study, we exposed chickens collected from local farmers in Mombasa, which is located at an elevation of approximately 50 m (lowland) in the Kenyan coast with an average temperature between 22°C and 35°C (Njarui et al., 2016), and from Naivasha, located at an elevation of approximately 1800 m (highland) with an average temperature of 8°C to 26°C (Ouko et al., 2017), to a short-term HS treatment (acute) and a repeated longer-term HS treatment (chronic) and analyzed the transcriptome response of skeletal and cardiac tissues using RNA sequencing. Exposure of chicken embryo to elevated temperature induces an adaptive response to HS at later stages in their life (Janke et al., 2004; Loyau et al., 2014; Loyau et al., 2015; Loyau et al., 2016; Fleming et al., 2017). It was hypothesized that chickens that were hatched at relatively higher temperature in the lowlands would respond to HS differently than the highland chickens that were hatched and raised at a lower temperature. Comparative transcriptome analysis by measuring global gene expression changes between the two will identify important genes and pathways that are critical for response to HS. We performed pairwise differential gene expression analysis between control and treatment groups at each time point and identified functional difference in response to HS between the tissues of the two chicken types. We then performed gene coexpression network (GCN) analysis by integrating the different differentially expressed gene (DEG) datasets generated from lowland and highland chickens to understand the overall response of the two ecotypes to HS.

MATERIALS AND METHODS

Experimental Design

The study involved two groups of chickens; one was collected from the lowland (low altitude) region and another from

the highland (high altitude) regions of Kenya. The lowland chickens were collected from local farmers in Mombasa (4°1′0″S, 39°35′24″E) (average temperature between 22°C and 35°C), while the highland chickens were obtained from KALRO (Kenyan Agricultural and Livestock Research Organization) in Naivasha (average temperature between 8°C and 26°C). A schematic of the experimental design is given in **Figure 1**. A total of 32 (n = 16 from each region), 5-month-old female chickens were used in this study. The HS experiments were conducted at the KOPIA (Korea Project for International Agriculture) Kenya center at Nairobi. The birds had *ad libitum* access to feed and water. After acclimating the birds to the local environment in the experimental cage for 3 days, the experiments were performed. The experiments were performed in a specially designed cage fitted with a temperature controller (**Supplementary Figure 1**). The HS group (n = 16) was exposed to high temperatures of 35°C for 8 h per day (9:00–17:00 h) and remained at 28°C to 30°C at all other times. The control group (n = 16) was maintained at 24°C during the entire experimental period. The short-term HS group (acute group) (n = 16, four per region, including the controls) were euthanized after 5 h of increased temperature exposure, and cardiac and skeletal muscle tissues were collected. The long-term HS group (C) (n = 16, four per ecotype, including the controls) were euthanized at the end of 3 days of cyclic HS, and cardiac and skeletal muscle tissues were collected. Rectal temperatures were measured at the beginning and end of the treatment period using a temperature probe. A total of 64 samples were collected; they were stored in RNAlater (Ambion, Texas, USA) and transported to the National Institute of Animal Science (South Korea) and stored at −80°C until further use.

RNA-Seq Analysis

Total RNA was isolated from 32 skeletal muscles and 32 cardiac muscles with RNeasy mini kit (Qiagen, USA) following the manufacturer's protocol. The purity and concentration of the isolated RNA were measured with NanoDrop ND-1000 UV-vis Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The integrity of the RNA was measured on Bioanalyzer 2100 system using RNA Nano 6000 Assay kit (Agilent Technologies, CA, USA), and only samples with a RIN (RNA integrity number) value greater than 8 were used for sequencing. cDNA libraries were generated using Illumina TruSeq® RNA sample preparation v2 kit (Illumina, San Diego, CA, USA) following processes previously described (Srikanth et al., 2017b). Quality of the individual libraries was accessed on Bioanalyzer 2100 system using DNA Nano 1000 Assay kit. Paired-end (PE) sequencing was performed on Illumina® HiSeq 2000 on four lanes (21, 12, 7, and 24 samples on lanes 1, 4, 5, and 6, respectively) of a single chip, blocking by treatment, tissue, and region. The sequencing was carried out by Macrogen (Seoul, South Korea). The raw reads are freely available at the NCBI (National Center for Biotechnology Information) SRA (Sequence Read Archive) database under accession number PRJNA557270. The quality of the raw reads was accessed using FastQC (version 0.11.5) (Andrews, 2010). Reads shorter

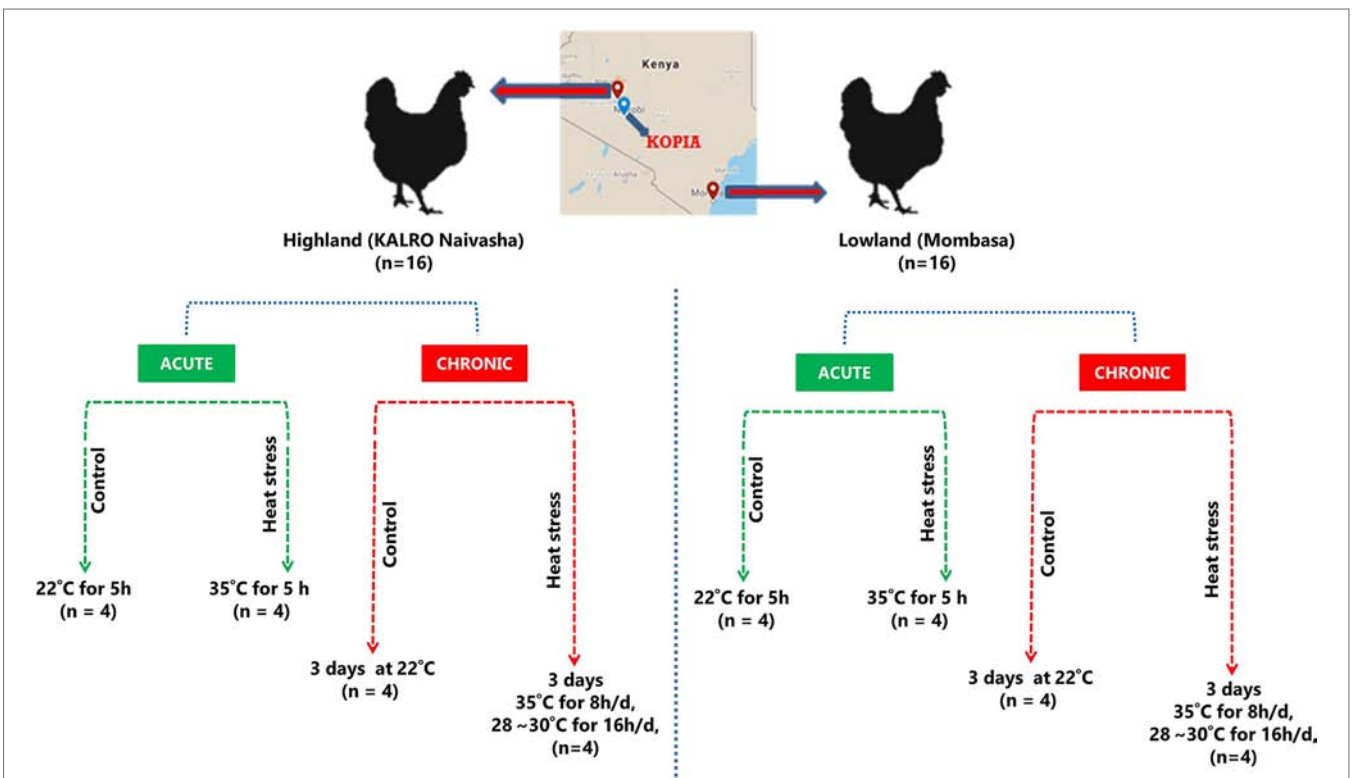


FIGURE 1 | Schematic experimental design.

than 80 base pairs (bp), low-quality bases, and adapters were removed using TRIMMOMATIC (version 0.36) (Bolger et al., 2014). All 100-bp reads were individually aligned to the chicken reference genome (*Gallus gallus* 5.0, release 94, Ensembl) using HISAT2 (version 2.0.5) (Kim et al., 2016) following methods previously described (Park et al., 2019). The DEGs were identified using CUFFLINKS (version 2.2.1) (Trapnell et al., 2012). The -G/–GTF flags was used to quantitate against reference transcript annotations. The expression of individual genes were measured as fragments per kilobase of exon per million, and DEGs (FDR < 0.05) were identified with CUFFDIFF. Eight DEG gene sets were generated (highland: control-acute, control-HS, HS-acute, HS-chronic; lowland: control-acute, control-HS, HS-acute, HS-chronic). Functional annotation and overrepresentative analyses were carried out using the web-based gene ontology (GO) clustering tool DAVID (Huang et al., 2008). The genes were annotated under the Biological Process, Molecular Function, and KEGG pathway terms. Significant terms (FDR <0.05 for KEGG and FDR <0.05 for GO terms) were plotted with ggplot2 package in R (version 3.4.1) (Team, 2013). The Venn diagrams were generated with a web-based tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

GCN Analysis and KEGG Pathway Mapping

GCN was constructed using the partial correlation coefficient with information theory (PCIT) algorithm (Reverter and Chan, 2008). We constructed two networks (a lowland-specific network and a highland-specific network), using DEGs (differentially expressed), in at least one of the four gene sets that were generated from the lowland or highland chicken groups. Only genes that had a partial correlation $|r|$ of ≥0.99 were included for network construction. The networks were visualized in CYTOSCAPE (version 3.4.1) (Shannon et al., 2003), analyzed with the NetworkAnalyzer plugin, and sorted according to degrees of connections. The genes in the network were then mapped to Kyoto Encylopedia of Genes and Genomes (KEGG) pathways (Ogata et al., 1999; Kanehisa and Goto, 2000) using the ClueGO plugin (Bindea et al., 2009).

Quantitative Reverse Transcriptase–Polymerase Chain Reaction Analysis

One microgram of each of the isolated RNA was reverse transcribed into cDNA with Oligo(dT) using SuperScript III™ first-strand system for reverse transcriptase–polymerase chain reaction (RT-PCR) (Invitrogen, CA, USA) in a final volume of 20 µl using the manufacturer's protocol. The resulting cDNAs were diluted 1:2, prior to their use for quantitative RT (qRT)–PCR analysis. The PCR reactions were carried out at a final volume of 10 µl containing 5 µl of Universal Master Mix containing dNTPs. $MgCl_2$, reaction buffer and AmpliTaq Gold DNA polymerase, 90 nM of primers (forward and reverse) and 250 nM of fluorescence-labeled TaqMan probe, and finally 2 µl of the cDNA. Amplifications were carried out on an ABI PRISM 7900HT Sequence Detection Systems (Applied Biosystems, CA, USA) with initial denaturing for 10 min at 95°C, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. All samples were amplified in triplicates. The data were analyzed with the SEQUENCE DETECTOR software (Applied Biosystems). All reagents used in the qRT-PCR analysis were procured from Life Technologies (Carlsbad, CA, USA). The absolute fold change was calculated after normalization with the chicken glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH), using the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008). All the primers used in the analyses are listed in **Table 1**.

RESULTS

All the chickens (n = 32) procured from the lowland Mombasa region and the highland Naivasha region were brought to the KOPIA center in Nairobi. After acclimating the birds to the local environment and the cage, the lowland and highland birds were randomly separated into four groups each (n = 4/group), comprising the acute stage groups: ALL (acute and lowland) and AHL (acute and highland); and the chronic stage groups; CLL (chronic and lowland) and CHL (chronic and highland). Each of these stages comprised a treatment group and a control groups (**Figure 1**).

Effect of HS on Rectal Temperature

Figure 2A shows the changes in the rectal temperatures of the animals before the start and at the end of the experiment. The maximum increases in rectal temperatures were in the AHL and the ALL HS groups. On an average, the rectal temperatures of AHL and ALL birds increased by 1.8°C and 1.6°C, respectively. Only minor changes in temperature were noted in the control groups.

Transcriptome Alignment and Mapping Statistics

We constructed 64 cDNA libraries from the cardiac and skeletal muscle tissues of the lowland and highland Kenyan chicken groups from the two experimental time points (acute and chronic). There were 1.33 billion (647 million in cardiac and 650 million in skeletal muscle), 100-bp PE reads corresponding

TABLE 1 | List of primers used for qRT-PCR validation of RNA-Seq results.

Gene	Primer sequencing (5′ –> 3′)
HSPH1	F –> TGAATTGGAAACTCAGGACCAGATG
	R –> CCTCACTGTTCTCTTGCTGGTTATT
SRGN	F –> GGCCGTGGTTCCAGCT
	R –> GCTCCTGGTACGTCTTCATCAG
ATRAID	F –> GTTGGACCTCAGCAACTGTTC
	R –> CGTCAGGTCCAGCACGAC
MT4	F –> CGGAGCTGCCGCAAGA
	R –> CCTTGGCACAGTTGTTGCA
PDK4	F –> TGACTGGTCGCATCCCAAGTAAG
	R –> GGAAGAATTTGCCTGTTTGGAGG
OTUD1	F –> GCTGTGTCCCTCTCCAAGATG
	R –> CACGCTTCTGTCCGTCTGT
GAPDH	F –> GGAAGAATTTGCCTGTTTGGAGG
	R –> TCGTCAAGCTTGTTTCCTGGTATGA

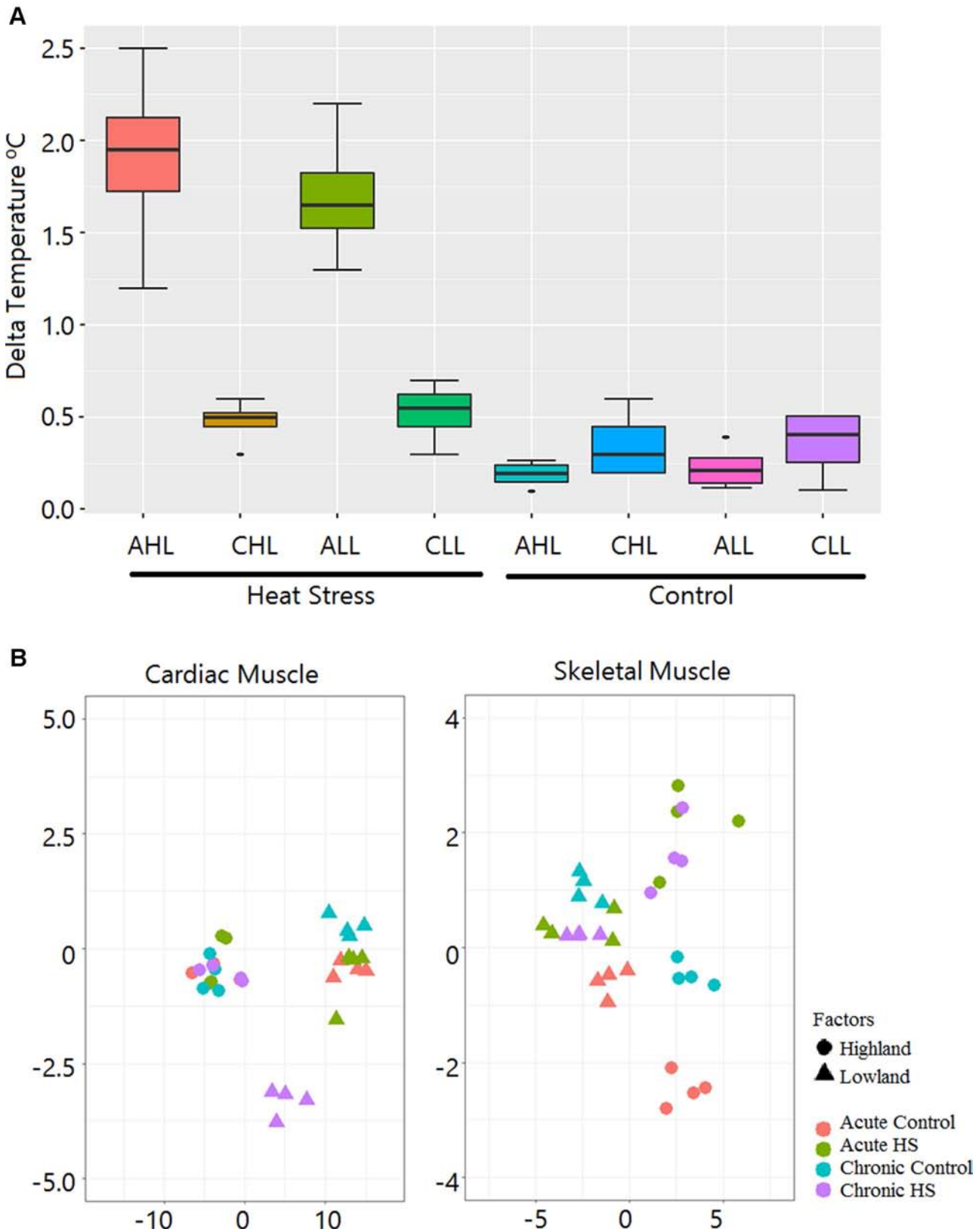


FIGURE 2 | Rectal temperature gain during HS treatment and sample relationship inferred from principal components analysis. **(A)** Box plot showing gain in rectal temperature due to HS treatment. AHL, acute highland; CHL, chronic highland; ALL, acute lowland; CLL, chronic lowland. **(B)** Principal components analysis showing that the maximum variation is due to differences between the ecotypes. Only a small percentage of the variation is due to the HS effect.

to an average of 1.65 Gb of sequence data per sample that were generated. After trimming for adapters and low-quality reads, 1.29 billion reads corresponding to 97.56% of total sequenced reads were used for downstream analysis. The reads were mapped to the chicken genome at an average alignment rate of 91.1% and 85.12% for cardiac and skeletal muscle tissues, respectively. A summary of the mapping statistics is given in **Supplementary File 3**. Principal components analysis showed that the maximum variation is due to difference between the ecotypes, and only a small percentage of the variation is due to HS effect (**Figure 2B**). A list of all the DEGs identified in this study is given in **Supplementary File 1**.

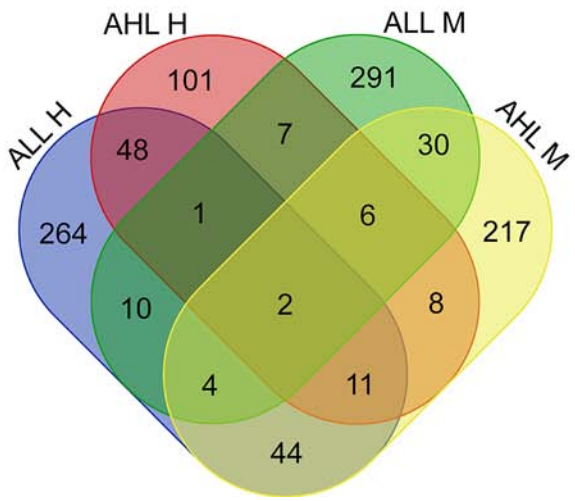
Effects of Acute HS on the Cardiac and Skeletal Muscle Transcriptome of the Highland and Lowland Chickens

Acute HS resulted in 351 and 322 genes in the skeletal muscle and 384 and 184 DEG in the cardiac tissues to be significantly differentially expressed (FDR <0.05) in the lowland and the highland, respectively (**Figure 3A**). Between the two ecotypes, 48 and 30 DEGs overlapped between the cardiac and skeletal muscle tissues, respectively (**Figure 3B**). Ten DEGs were commonly differentially expressed between the two tissues in the lowland chickens; eight DEGs differed between the two tissues in the highland chickens (**Figure 3B**). Only two DEGs were found in all four contrasts. These were heat shock protein (HSP) family A (Hsp70) member 8 (HSPA8) and HSP family B (small) member 7 (HSPB7). GO enrichment analysis (**Figure 4A**) of the up-regulated DEGs showed that at the acute stage in the cardiac tissue of lowland chicken, the most significant terms (Q <0.05) enriched were "signal transduction," "immune response "activation of MAPKK activity," "fatty acid binding," "response to unfolded protein," and "apoptosis," while in the highland chickens "signal transduction," "serine-type endopeptidase activity," "inflammatory response," "immune response," and "apoptosis" were the most significantly enriched terms. In the skeletal muscle of lowland chickens, the GO terms enriched were "response to heat," "protein folding," "apoptosis," "oxido-reductase activity," "activation of MAPKK activity," and "positive regulation of ERK1 and ERK2 cascade," while in the highland chickens, "signal transduction," "protein folding," "apoptosis," "oxido-reductase activity," "activation of MAPKK activity," and "immune response." Among the down-regulated DEGs (**Figure 4C**) "protein kinase inhibitor activity,"

A

DEGs	ALL M	CLL M	ALL H	CLL H	AHL M	CHL M	AHL H	CHL H
Up	259	82	136	61	267	91	115	114
Down	92	60	248	111	55	89	69	56
Total	351	142	384	172	322	180	184	170

B



C

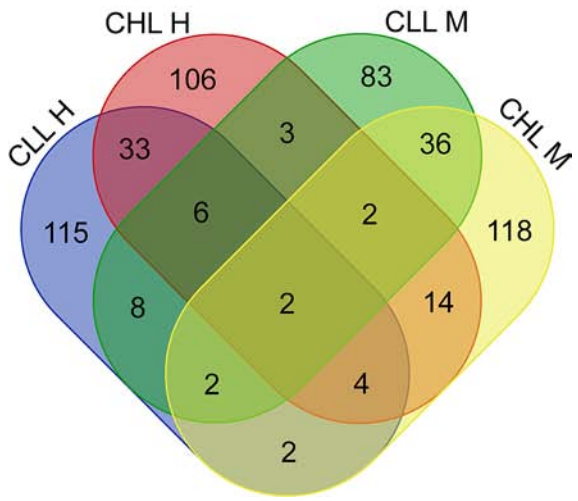


FIGURE 3 | Genes differentially expressed relative to control group under different contrast. **(A)** Number of genes that were up-regulated or down-regulated after HS treatment. **(B)** Venn diagram showing common and unique genes differentially expressed during the acute stage treatment. **(C)** Venn diagram showing common and unique genes differentially expressed during the chronic stage treatment. M and H, skeletal muscle and cardiac muscle, respectively; AHL, acute highland; CHL, chronic highland; ALL, acute lowland; CLL, chronic lowland.

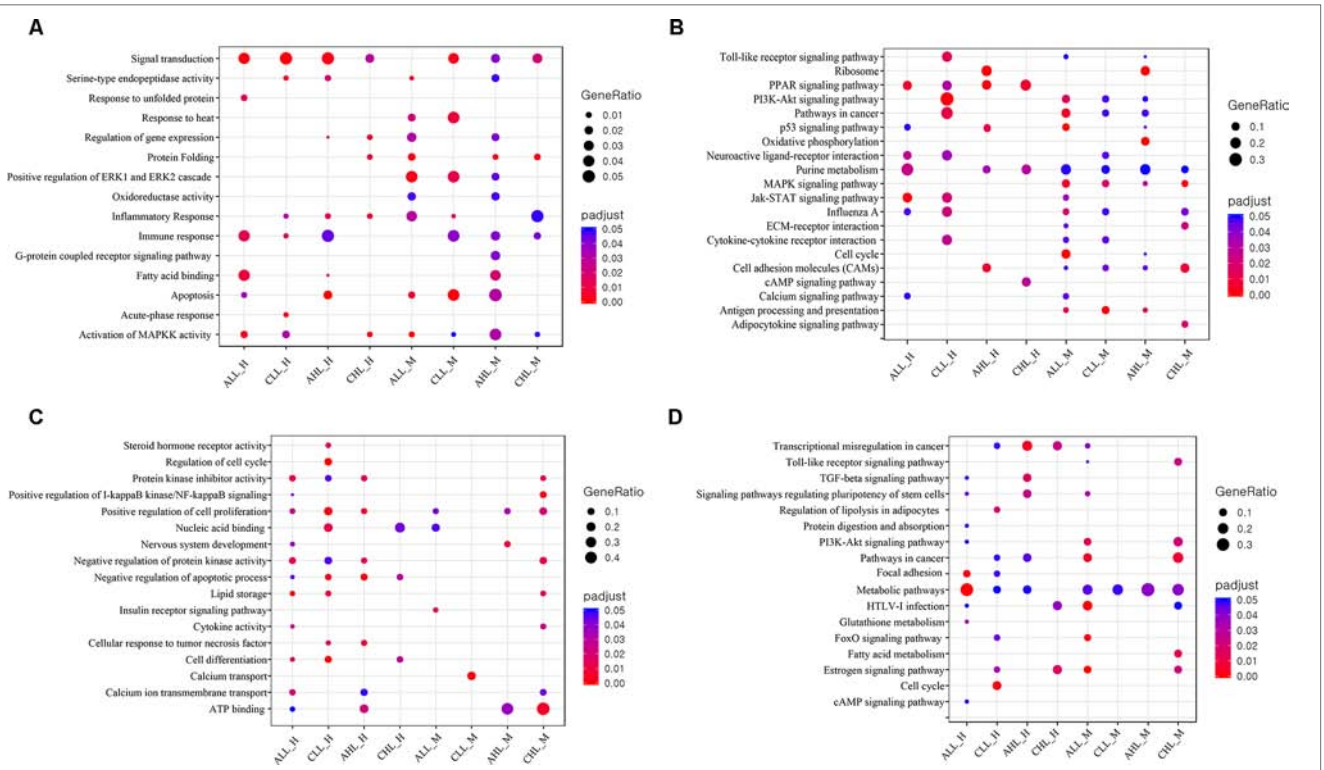


FIGURE 4 | GO and KEGG pathway enrichment analysis **(A)** Dot plot shows the up-regulated GO terms (FDR <0.05) of biological processes and molecular functions identified using DAVID to be enriched under the different contrasts. **(B)** Dot plot shows the up-regulated KEGG pathways (FDR <0.1) enriched for different contrasts. The size of the dot is based on gene count enriched in the pathway, and the color of the dot shows the pathway enrichment significance. **(C)** Dot plot shows the down-regulated GO terms (FDR <0.05) of biological processes and molecular functions identified using DAVID to be enriched under the different contrasts. **(D)** Dot plot shows the down-regulated KEGG pathways (FDR <0.1) enriched for different contrasts. The size of the dot is based on gene count enriched in the pathway, and the color of the dot shows the pathway enrichment significance.

"positive regulation of cell proliferation," "negative regulation of apoptotic process," lipid storage," and "cell differentiation" were significantly down-regulated in cardiac tissues of acute stage lowland chickens, while "positive regulation of cell proliferation," "negative regulation of apoptotic process," "cellular response to tumor necrosis factor," and "ATP binding" were down-regulated in the highland chicken. In the skeletal muscle, "insulin receptor signaling" and "calcium transport" were the most significantly enriched terms in the acute stage in the lowland chickens, while "ATP binding," "nervous system development," and "positive regulation of cell proliferation" were significantly enriched in the highland chickens. The enrichment analysis showed that several biological processes were regulated in more than one situation.

KEGG pathway enrichment analysis of up-regulated DEGs (**Figure 4B**) showed that in the cardiac muscles of the lowland chickens the most enriched pathways (FDR <0.05) were "Jak-STAT signaling pathway," "PPAR signaling pathway," "purine metabolism," "neuroactive ligand–receptor interaction," and "p53 signaling," while in the skeletal muscle tissue "cell cycle," "MAPK signaling pathway," "pathways in cancer," "antigen processing and presentation," and "p53 signaling pathway" were enriched. In the cardiac tissues of highland chickens, "PPAR signaling pathway," "cell adhesion molecules," and "p53 signaling pathway" and in the skeletal muscle "MAPK signaling," "oxidative phosphorylation," "PPAR signaling pathway," "ribosome," and "purine metabolism" were enriched. Among the down-regulated DEGs (**Figure 4D**) in the cardiac tissue of lowland chickens, "metabolic pathways" and "focal adhesion" were significantly down-regulated, while "estrogen signaling pathway," "HTLV-I infection," and "FoxO signaling pathway" were down-regulated in skeletal muscle. In the cardiac muscle of highland chickens, "transcriptional misregulation in cancer" and "TGF-beta signaling pathway" were significantly down-regulated, while "metabolic pathways" was down-regulated in the skeletal muscle.

Effects of Chronic HS on the Cardiac and Skeletal Muscle Transcriptome of the Highland and Lowland Chickens

Under chronic HS, 142 and 172 DEGs were found in the skeletal and cardiac tissues of the lowland chickens, while 180 and 170 DEGs were found in the skeletal and cardiac tissues of the highland chickens (**Figure 3A**). Between the two chicken ecotypes, 33 DEGs were common between the cardiac and 36 DEGs were common between the skeletal muscle tissues (**Figure 3C**). Eight DEGs were common between the two tissues in the lowland chicken; 14 DEGs were common between the two tissues in the highland chicken (**Figure 3C**).

Only two DEGs were common in all the four contrasts; these were the HSP family A (HSP70) member 8 (HSPA8) and fatty acid–binding protein 4 (FABP4) (**Figure 3C**), these two genes were previously found be differentially expressed in the hypothalamus of a meat-type chicken (Sun et al., 2015a). GO enrichment analysis of the up-regulated DEGs (**Figure 4A**) showed that in the lowland chickens "signal transduction," "acute-phase response," "immune response," "inflammatory response," and "serine-type endopeptidase activity" were enriched in the cardiac tissue, while "apoptosis," "response to heat," "positive regulation of ERK1 and ERK2 cascade," and "immune response" were enriched in skeletal muscle. In the highland chicken, "regulation of gene expression," "protein folding," "inflammatory response," and activation of MAPKK activity" were enriched in the cardiac tissues, while "inflammatory response," "protein folding," and "signal transduction" were enriched in skeletal muscle tissues. Among the down-regulated DEGs (**Figure 4C**), "regulation of cell cycle," "positive regulation of cell proliferation," "nucleic acid binding," "lipid storage," and "cell differentiation" were enriched in the cardiac muscle of lowland chickens, while "calcium transport" was enriched in the skeletal muscle. In the highland chicken, "negative regulation of apoptotic process," "cell differentiation," and "nucleic acid binding" were enriched in the cardiac muscle, while "ATP binding," "positive regulation of I-kappaB kinase/NF-kappa signaling," and "protein kinase inhibitor" were enriched in the skeletal muscle.

KEGG pathway enrichment analysis of up-regulated DEGs (**Figure 4B**) showed that in lowland chickens "PI3K-Akt signaling," "pathways in cancer," "Jak-STAT signaling," and "Toll-like receptor signaling pathway" were enriched in the cardiac tissue, and "MAPK signaling pathway," "antigen processing and presentation," and "cell adhesion molecules" were enriched in the skeletal muscle. In the highland chickens, "PPAR signaling pathway," "purine metabolism," and "cAMP signaling" pathways were enriched in the cardiac and "MAPK signaling pathway," "ECM–receptor interaction," and "cell adhesion molecules" pathways were enriched in the skeletal muscle. "Cell cycle," "regulation of lipolysis in adipocytes," and "estrogen signaling" pathways were down-regulated (**Figure 4D**) in the cardiac muscle, while "metabolic pathway" was down-regulated in the skeletal muscle in lowland chicken. In the highland chicken, "estrogen signaling pathway," "HTVL-I infection," and "transcriptional misregulation in cancer" were down-regulated (**Figure 4D**) in cardiac muscle, while "pathways in cancer," "fatty acid metabolism," and "metabolic" pathways were down-regulated in the skeletal muscle.

Integration of Cardiac and Skeletal Muscle DEGs to Understand the Overall Response of the Highland and Lowland Chickens to HS

We integrated the DEGs identified in the acute and chronic stages in the cardiac and skeletal muscle tissues of the highland (**Figure 5A**) and lowland (**Figure 5C**) chickens through a GCN constructed based on PCIT (Reverter and Chan, 2008). The highland GCN comprised of 75 nodes (genes) and 244 edges (connections) (**Figure 5A**). KEGG pathway enrichment analysis of the genes in the network showed that four pathways comprising 28 of the 77 genes in the networks were enriched; these included "PPAR signaling pathway," "protein processing in endoplasmic reticulum," "MAPK signaling pathway," and "p53 signaling pathway' (**Figure 5B**). The lowland gene network comprised 77 nodes (genes) and 270 edges (connections) (**Figure 5C**). Three KEGG pathways comprising 25 of the 75 genes in the network were found to be enriched; these included "p53 signaling," "steroid biosynthesis," and "PPAR signaling" pathways (**Figure 5D**). The networks were sorted according to degree (number of edges incident to the node [genes]). A list of all the genes in the network and their degree is given in **Supplementary File 2**. Genes with the maximum connections (degrees) in the lowland network included cyclin-B2 (*CCNB2*), crumbs homolog 2 (*Crb2*), carbohydrate sulfotransferease 9 (*CHST9*), sestrin-1 (*SESN1*), and nuclear receptor subfamily 4 group A member 3 (*NR4A3*), while in the highland coexpression network COMM domain containing 4 (*COMMD4*), tetratricopeptide repeat domain containing protein 32 (*TTC32*), H1 histone family member 0 (*H1F0*), acylphosphatase 1 (*ACYP1*), and ribosomal protein S28 (*RPS28*) had the highest degree. Comparison of the two networks revealed that 30 genes were shared between the two networks (**Supplementary File 2**); among these were genes involved in PPAR signaling pathway (*PLIN1*, *SCD*, *FABP4*, *FABP1*, and *DBI*), p53 signaling pathway (*CDK1*, *TP53I3*, *GADD45B*, *SESN1*, *GTSE1*), and MAP Kinase signaling pathway (*DUSP5* and *DUSP8*).

Validation of RNA-Seq Results

Out of 30 genes that overlapped between the two coexpression networks, six genes were randomly chosen for validation by qRT-PCR analysis. **Figure 6** shows the PCR quantification of the *OTUD1*, *HSPH1*, *PDK4*, *ATRAID*, *SRGN*, and *MT4* genes. The results broadly showed a similar expression profile between the RNA-Seq and qRT-PCR. A correlation of 0.86 was observed between the RNA-Seq and qRT-PCR log2 fold-change results (**Figure 6**).

DISCUSSION

High ambient temperatures affect the production and reproduction rates in animals (Srikanth et al., 2017a). Studies have highlighted the deleterious effect of HS on physiological (Altan et al., 2003; Mujahid et al., 2007), biochemical (Xie et al., 2015), and immune capacity of chickens (Park et al., 2019; Te Pas et al., 2019) (Altan et al., 2003; Mujahid et al., 2006; Mujahid et al., 2007; Huang et al., 2015; Xie et al., 2015; Park et al., 2019; Te Pas et al., 2019). Fast-growing, commercial chickens, artificially selected and raised under controlled environment, are very sensitive to HS and might not have the genetic potential to develop thermal tolerance, limiting their potential for rearing in developing countries (Coble et al., 2014; Lan et al., 2016; Fleming et al., 2017). Native breeds and village ecotypes that have been under environmental challenges such

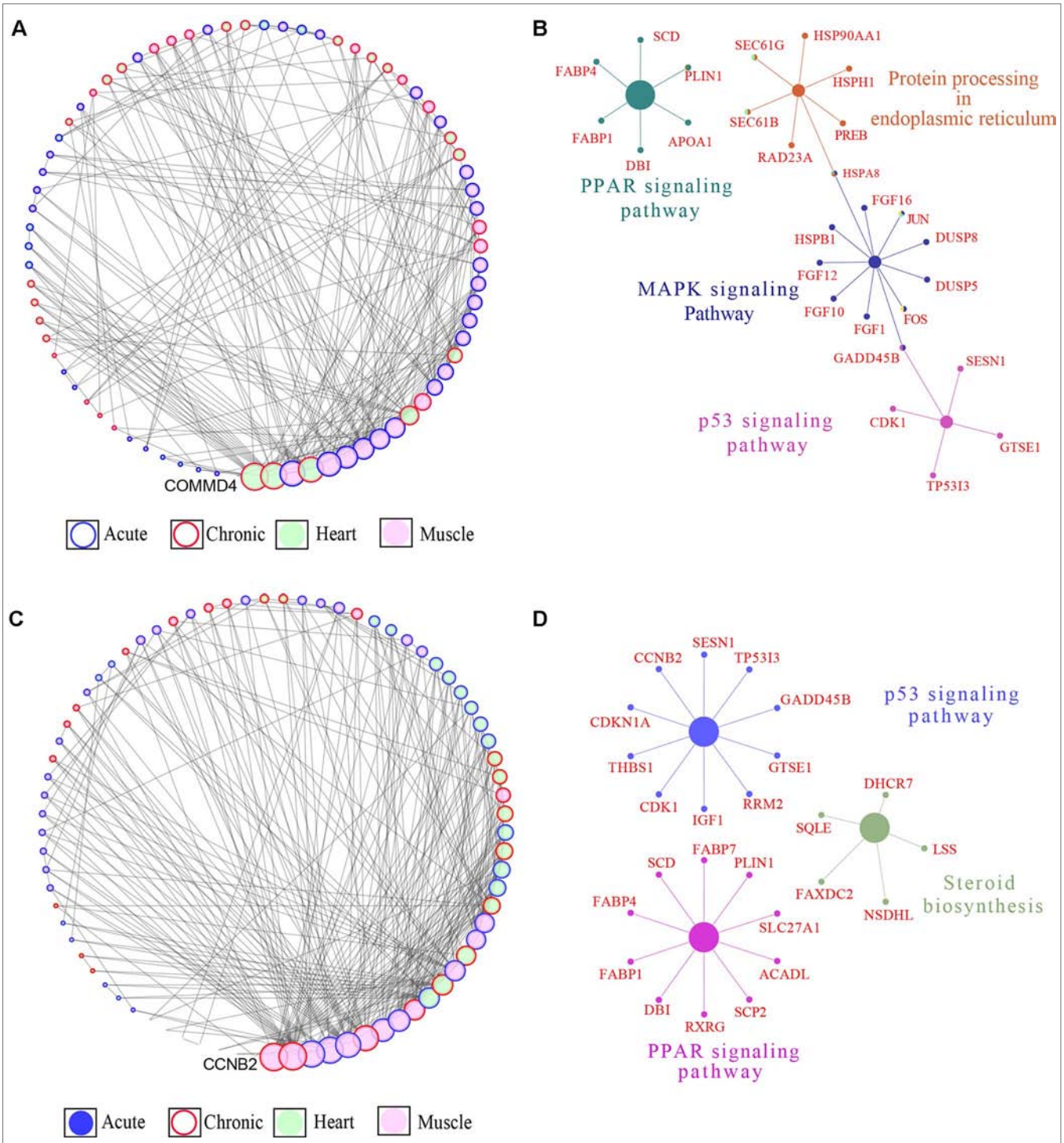


FIGURE 5 | Gene coexpression network (GCN) and pathway enrichment analysis integrated for the skeletal and cardiac muscle DEGs. **(A)** Degree sorted network of DEGs in at least one contrast in the highland chickens. The nodes are genes, and the edges are based on correlation coefficients. Only genes with a partial correlation | *r* | of ≥0.99 were included in network. Node color denotes the tissue type in which the gene expression was the highest, while node border denotes the stage at which the gene expression was the highest. **(B)** KEGG pathway networks in which all the genes in the highland GCN network were enriched. **(C)** Degree sorted network of DEG in at least one contrast in the lowland chickens. The nodes are genes, and the edges are based on correlation coefficients. Only genes with a partial correlation | *r* | of ≥0.99 were included in network. Node color denotes the tissue type in which the gene expression was the highest, while node border denotes the stage at which the gene expression was the highest. **(D)** KEGG pathway networks in which all the genes in the lowland GCN network were enriched.

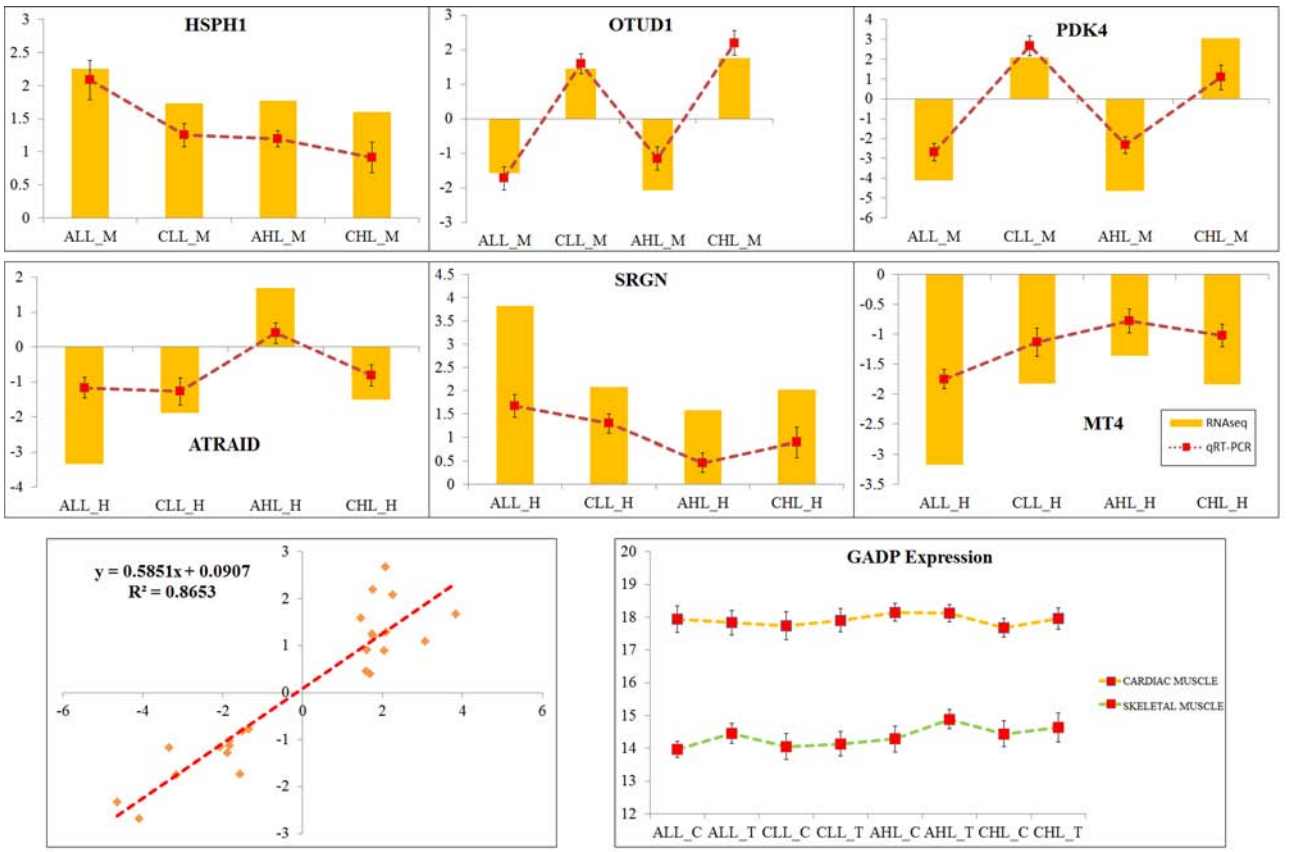


FIGURE 6 | Validation of RNA sequencing results with qRT-PCR analysis. The real-time PCR analyses of HSPH1, OTUD1, PDK4, ATRAID, SRGN, and MT4 were performed on all the samples that were used in the RNA-Seq analysis. GADPH, which showed constant expression in all samples, was used as a normalization gene. Log2 fold changes identified in the RNA-Seq analysis were plotted as bar, and the fold changes through qRT-PCR were plotted as line graph. A scatter plot of RNA-Seq Log2 fold change and the qRT-PCR fold change were plotted, and the correlation (r^2) between the two methods was identified.

as high ambient temperature over multiple generations might have developed thermal tolerance due to adaptation to local conditions (Clarke, 2003; Chen et al., 2009; Seebacher, 2009; Nardone et al., 2010; Lawrence and Wall, 2014; Porto-Neto et al., 2014); examining such native breeds will provide us with genetic information needed to mitigate the impact of HS. In this study, we explored the transcriptomic response of Kenyan chicken ecotypes collected from two different environmental regions. The cardiac muscle was chosen due to its central role in heat dissipation through blood circulation (Zhang et al., 2017), and skeletal muscle was chosen due to its susceptibility to HS-induced oxidative damage and damages to membrane integrity (Sandercock et al., 2001; Mujahid et al., 2006), which affects meat quality.

Difference in HS Response Between Acute and Chronic Stages

Considerably higher changes in rectal temperature were noted in the acute group birds, suggesting difficulties in maintaining the core body temperature in response to sudden increase in temperature, compared to the chronic group birds, which were able to regulate their body temperature better at the end of the experimental period. This may be due to acclimation for cyclic HS. The highest change in rectal temperature was observed in the highland chickens of the acute group. These chickens were less adapted to HS as compared to lowland chickens. Therefore, it may be assumed that the response of highland chickens to HS is less robust than the response of lowland chickens. While response to acute HS is under homeostatic regulation (reflex-responsive regulation), the response to chronic HS is under homeorhetic regulation, i.e., metabolic regulation through endocrinal hormones (Collier et al., 2018). Studies have shown that in chickens acute HS, i.e., shock due to sudden change in ambient temperature, is more stressful (Lan et al., 2016) than cyclic HS (Coble et al., 2014). In the lowland chicken, in both cardiac and skeletal muscles, the acute group had 2.2 to 2.5 times more DEGs than the chronic group; however, in the highland chickens, while there were 1.7 times more DEGs in the acute stage of skeletal muscle tissues, a similar number of DEGs were found at both stages in the cardiac tissue (**Figure 3A**). This suggests a difference in heat sensitivity in the cardiac tissues between the lowland and highland chickens. The overall increased number of DEGs in the lowland chickens and the comparatively lesser

change in rectal temperature (**Figure 2**) suggest a considerably stronger response to HS than the highland chickens. This difference may denote the difference in acclimatization (rate) between highland and lowland chickens due to a difference in adaptation to HS.

Between the different contrasts in the acute and chronic stages, very few genes overlapped. Overall only two genes in cardiac (HSPA8 and HSPB7) and two genes in skeletal muscle (HSPA8 and FABP4) overlapped. Thus, we conclude that acclimatization leads to a complete change in response to HS. The HSPA8 gene (up-regulated in all contrast) is a member of the HSP70 family of molecular chaperones and is known to play an important role in directing correct folding of newly synthesized proteins and in the destruction of non-reversibly denatured proteins (Hartl, 1996; Iwamoto et al., 2005). HSPA8 has been found to be up-regulated under HS in chickens (Sun et al., 2015b; Wang et al., 2015) and could serve as a good biomarker. HSPB7 is a member of the small HSP (sHSP) family, whose expression is restricted to skeletal and cardiac muscles (Bonomini et al., 2018). HSPB7 functions in protecting cells from protein aggregation (Vos et al., 2010) and is required for maintaining muscle integrity (Juo et al., 2016). HS causes aggregation of denatured proteins (Srikanth et al., 2017a). HSPB7 was found to be significantly up-regulated in lowland chicken's cardiac (acute and chronic) and skeletal muscle (acute), while the effect was opposite (down-regulated) in the highland chickens. This could indicate that the lowland chickens might be able to counteract the effects of HS-induced protein aggregation better than the highland chickens. We also observed stage-specific difference in HS response (**Figure 4B**). The p53 pathway, which is essential for DNA damage repair, initiation of cell cycle arrest, and cell apoptosis (Rappold et al., 2001; Harris and Levine, 2005), was enriched in both cardiac and skeletal muscles in the acute stage, suggesting that acute HS might have an inhibiting effect on cell cycle; similar observation was noted in liver of heat-stressed broilers (Jastrebski et al., 2017). The PPAR signaling pathway, which is required for energy metabolism (Wang, 2010) and regulating the oxidative stress–induced inflammatory response (Kim et al., 2017), was found to be enriched under chronic HS in three of the four contrasts (**Figure 4B**). Prolonged exposure to heat can cause considerable oxidative stress damage in chickens (Azad et al., 2010; Akbarian et al., 2016). The enrichment of PPAR signaling in the chronic HS group could be indicative of HS-induced reactive oxygen species (ROS) accumulation and oxidative stress. While this may relate to the inflammatory response, it may also just indicate differences between tissues.

Difference in HS Response Between Cardiac and Skeletal Muscle

The overlapping of very few genes between cardiac and skeletal muscle within and between the two ecotypes (lowland and highland) (**Figures 3B, C**) shows that not only is there a difference in HS response between the tissues, but there is also an ecotype-specific difference in HS response, suggesting differences in HS adaptation. This was indicated by the enrichment of the PPAR signaling pathway in the cardiac and the MAPK signaling pathway in the skeletal muscle. While PPAR signaling pathway regulates energy metabolism and regulation of the oxidative stress response (Wang, 2010; Kim et al., 2017), the MAPK signaling is required for activating programmed cell death (Pearson et al., 2001). The activation of members of the PPAR signaling pathway in cardiac muscle might be due to the requirement of considerable energy for pumping blood to dissipate the accumulated heat or to alleviate oxidative stress, while the enrichment of MAPK signaling genes could be indicative of cellular damage in the skeletal muscle and the triggering of apoptosis.

Difference in HS Response Between Lowland and Highland Chicken

The DEGs generated under different contrasts were integrated into lowland and highland gene correlation networks (GCNs) to study their overall response to HS. The GCNs were then sorted for number of edges, denoting connections between nodes (genes), to identify hub genes in each ecotype. The top 5 hub genes in the lowland GCN all function in regulating cell cycle, cell signaling, cell division, or in DNA repair mechanism. Cyclin B2 (CCNB2), which is an important regulator of cell mitosis (Brandeis et al., 1998), was significantly up-regulated in lowland skeletal muscle at the chronic stage. CCNB2 was previously found to be significantly down-regulated in the fast-growing ROSS 708 broilers compared to the slow-growing Illinois broiler under HS and was suggested to be indicative of the reduction in cell cycle activity in the Ross broilers (Zhang et al., 2017). The Crb2, a BRCT protein, is a cell cycle checkpoint mediator that is essential for cellular response to DNA damage and repair (Kilkenny et al., 2008). Crb2 was found to be significantly up-regulated in the skeletal muscle at acute and chronic stages. Carbohydrate *N*-acetylgalactosamine-4-O-sulfotransferase 9 (CHST9), a member of the N-acetylgalactosamine-4-O-sulfotransferase family that catalyzes the transfer of sulfate to position 4 of nonreducing terminal GalNAc residues, is implicated in cellular signaling events (Xia et al., 2000; Baenziger, 2003; Zhao et al., 2010). The expression of CHST9 was found to be up-regulated in the acute stage in skeletal muscle. SESN1, which was found to be critical for prolonging the life span of *Caenorhabditis elegans* by preventing muscle degeneration, is required for ROS clearance and plays a key role in defense against HS (Yang et al., 2013). The expression of SESN1 was found to be up-regulated in the skeletal muscle in the acute stage. The NR4A3 is a nuclear orphan receptor and a member of the Nur77 family. The NR4A3 activates several genes that are critical for regulating cell cycle, inflammation, and DNA repair (Wenzl et al., 2015). NR4A3 expression was found to up-regulated in the skeletal muscle of lowland chicken. The differential expression of these key hub genes, which had significant expression correlation with 91 other genes in the network (degrees), suggests that the lowland chickens were affected by HS, and they responded robustly by activating, cell cycle checkpoints, cell cycle arrestors, ROS clearance, and DNA damage and repair mechanisms.

The top hub genes in the highland coexpression network (**Figure 5A**), identified by the degrees of connections, were *COMMD4*, *TTC32*, *H1F0*, and *ACYP1*. COMM (Copper metabolism gene MURR1) domain containing 4 (COMMD4) is an inhibitor of TNF (tumor necrosis factor)–induced NF-κB (nuclear factor κB) (Burstein et al., 2005). Activated NF-κB regulates the expression of several genes that controls cell proliferation, apoptosis, and inflammation (Liu et al., 2016). The expression of COMMD4 was found to be significantly down-regulated in the cardiac (acute and chronic stages) and skeletal muscle (acute stage) of highland chickens, possibly indicating the activation of NF-κB transcription regulatory factor. TTC32 was found to be up-regulated in the cardiac and skeletal muscle tissues at chronic and acute stages, respectively. The function of TTC32 is unknown; however, a number of TPR (tetratricopeptide repeat domain) interact with HSP family HSP70, HSP70, and HSP90 and are required for regulation of protein folding and transport (Ballinger et al., 1999), and HSP90AA1, a member of the HSP90 family, was significantly up-regulated in the skeletal muscle in the acute stage. H1F0 is involved in apoptotic DNA fragmentation (Wang et al., 2018) and contributes to labeling DNA damage (Keck et al., 2018). The expression of H1F0 was found to be up-regulated at both stages in the skeletal muscle and at the chronic stage in cardiac tissue in the highland chickens. The expression of ACYP1, an isoform of ACYP, which can induce apoptosis and is involved in ion transport (Degl'innocenti et al., 2004; Degl'innocenti et al., 2019), was found to be significantly elevated in the cardiac (chronic) and skeletal muscle (acute), while the expression of ACYP2 (another isoform of ACYP) was found to be elevated in the cardiac (Acute) and skeletal muscle (Chronic). This might indicate the activation of apoptosis due to DNA or cellular damage.

Pathway enrichment analysis of the genes in the GCNs (**Figures 5B, D**) showed that the p53 signaling and PPAR signaling pathways were enriched in both the lowland and highland networks. While p53 signaling pathway plays a pivotal role in cell death and cell survival by activating genes that induce cell cycle regulation, DNA repair, and cell death (Vogelstein et al., 2000; Zhang et al., 2010), PPAR signaling is critical for energy homeostasis (Wang, 2010), and considerable energy is spent in maintaining body temperature and dissipating heat, under hyperthermic condition. Moreover, there is a transcriptional dependence on PPAR for heat shock response (Vallanat et al., 2010). The steroid biosynthetic pathway was enriched only in the lowland network (Cook et al., 2015). Genes in this pathway can modulate the activities of RORγ (retinoic acid–related orphan receptors), which can regulate apoptosis (Kurebayashi et al., 2000). The MAPK signaling pathway and protein processing in endoplasmic reticulum (PP-ER) pathway were enriched only in the highland network (**Figure 5B**). HS is proteotoxic, and denatured proteins can become cytotoxic by forming aggregates (Srikanth et al., 2017a). Cell responds to this by activating the PP-ER pathway (Harding et al., 1999), which increases the protein-folding capacity in the endoplasmic reticulum and also activates apoptosis (programmed cell death) (Welihinda et al., 1999). Heat shock is known to activate several members of the MAPK family, which constitutes serine/threonine kinases that play a crucial role in transmitting signals required for cell growth, differentiation, and apoptosis (Pearson et al., 2001; Gorostizaga et al., 2005). The enrichment of multiple apoptosis activation pathways and hub genes that are proapoptotic factors in the highland network suggests that considerable cellular damage has taken place in the highland chickens

CONCLUSION

This study examined the transcriptome response to HS of two IC ecotypes from the lowlands and highlands of Kenya. Rectal temperature measurements and RNA-Seq analysis revealed that comparing the responses of acute HS and chronic HS indicated acclimatization of both lowland and highland chickens in such a short period. Furthermore, the response to HS is tissue and stage specific. The GCN analysis showed that the hub genes identified in the lowland chickens were cell cycle arrestors and DNA repair genes, while the highland hub genes were apoptotic and oxidative stress–responsive genes. These results lead us to conclude that, although both the ecotypes experienced HS, the lowland chickens responded more robustly than the highland chickens and might have a higher tolerance to HS than the highland chickens. This better acclimatization may be due to previous adaptation to higher temperatures in the lowland environment. This study extends our understanding of the HS response of chickens, and the genes and pathways identified could serve as a foundation for improving thermal tolerance in chickens.

DATA AVAILABILITY STATEMENT

The raw reads are available at the NCBI (National Center for Biotechnology Information) SRA (Sequence Read Archive) database under accession number PRJNA557270.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee, National Institute of Animal Science, South Korea.

AUTHOR CONTRIBUTIONS

MB, MP, and J-EP conceived the project. SK and DL collected the samples. KS, MB, HK, and WP performed the experiments. MB and J-EP secured the funding for the project. KS, J-MK, and WP analyzed the data. KS and J-EP interpreted the results and drafted the manuscript. J-MK and MP edited the manuscript.

FUNDING

This study was carried out with the support of "Investigation of expression profiles and genetic network related to heat stress for chicken" (project PJ01122101) and "Experimental metadata generation for sharing genome and metagenome data of Korean

and African chickens" (project PJ01275601), Rural Development Administration (RDA), Republic of Korea. KS and HK were supported by a 2019 RDA Fellowship Program of National Institute of Animal Science, Rural Development Administration, Republic of Korea.

ACKNOWLEDGMENTS

The authors thank the staff of International Livestock Research Institute (ILRI, Kenya), Kenyan Agricultural and Livestock Research Organization (KALRO, Naivasha), and Korea Project for International Agriculture (KOPIA, Kenya) for all the help made available for successful completion of this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00993/full#supplementary-material

REFERENCES

Akbarian, A., Michiels, J., Degroote, J., Majdeddin, M., Golian, A., and De Smet, S. (2016). Association between heat stress and oxidative stress in poultry; mitochondrial dysfunction and dietary interventions with phytochemicals. *J. Anim. Sci. Biotechnol.* 7, 37. doi: 10.1186/s40104-016-0097-5

Altan, Ö., Pabuçcuoğlu, A., Altan, A., Konyalioğlu, S., and Bayraktar, H. (2003). Effect of heat stress on oxidative stress, lipid peroxidation and some stress parameters in broilers. *Br. Poult. Sci.* 44, 545–550. doi: 10.1080/00071660310001618334

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Azad, M., Kikusato, M., Maekawa, T., Shirakawa, H., and Toyomizu, M. (2010). Metabolic characteristics and oxidative damage to skeletal muscle in broiler chickens exposed to chronic heat stress. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 155, 401–406. doi: 10.1016/j.cbpa.2009.12.011

Baenziger, J. (2003). *Glycoprotein hormone GalNAc-4-sulphotransferase*. London: Portland Press Limited. doi: 10.1042/bst0310326

Ballinger, C. A., Connell, P., Wu, Y., Hu, Z., Thompson, L. J., Yin, L.-Y., et al. (1999). Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Mol. Cell. Biol.* 19, 4535–4545. doi: 10.1128/MCB.19.6.4535

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi: 10.1093/bioinformatics/btp101

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bonomini, F., Favero, G., Trapletti, V., and Rezzani, R., (2018). "*HSP70 in human diseases and disorders*.," in *Heat shock proteins in cardiovascular diseases: from bench to bedside*, 223-245. (Switzerland: Springer). doi: 10.1007/978-3-319-89551-2_11

Brandeis, M., Rosewell, I., Carrington, M., Crompton, T., Jacobs, M. A., Kirk, J., et al. (1998). Cyclin B2-null mice develop normally and are fertile whereas cyclin B1-null mice die in utero. *Proc. Natl. Acad. Sci.* 95, 4344–4349. doi: 10.1073/pnas.95.8.4344

Burstein, E., Hoberg, J. E., Wilkinson, A. S., Rumble, J. M., Csomos, R. A., Komarck, C. M., et al. (2005). COMMD proteins, a novel family of structural and functional homologs of MURR1. *J. Biol. Chem.* 280, 22222–22232. doi: 10.1074/jbc.M501928200

Canario, L., Mignon-Grasteau, S., Dupont-Nivet, M., and Phocas, F. (2013). Genetics of behavioural adaptation of livestock to farming conditions. *Animal* 7, 357–377. doi: 10.1017/S1751731112001978

Chen, C.-F., Gourichon, D., Huang, N.-Z., Lee, Y.-P., Bordas, A., and Tixier-Boichard, M. (2009). Performance comparison of dwarf laying hens segregating for the naked neck gene in temperate and subtropical environments. *Genet. Sel. Evol.* 41, 13. doi: 10.1186/1297-9686-41-13

Chen, F., Yu, Y., Haigh, S., Johnson, J., Lucas, R., Stepp, D. W., et al. (2014). Regulation of NADPH oxidase 5 by protein kinase C isoforms. *PloS ONE* 9, e88405. doi: 10.1371/journal.pone.0088405

Ciscar, J.-C., Iglesias, A., Feyen, L., Szabó, L., Van Regemorter, D., Amelung, B., et al. (2011). Physical and economic consequences of climate change in Europe. *Proc. Natl. Acad. Sci.* 108, 2678–2683. doi: 10.1073/pnas.1011612108

Clarke, A. (2003). Costs and consequences of evolutionary temperature adaptation. *Trends Ecol. Evol.* 18, 573–581. doi: 10.1016/j.tree.2003.08.007

Coble, D. J., Fleming, D., Persia, M. E., Ashwell, C. M., Rothschild, M. F., Schmidt, C. J., et al. (2014). RNA-Seq analysis of broiler liver transcriptome reveals novel responses to high ambient temperature. *BMC Genomics* 15, 1084. doi: 10.1186/1471-2164-15-1084

Collier, R. J., Baumgard, L. H., Zimbelman, R. B., and Xiao, Y. (2018). Heat stress: physiology of acclimation and adaptation. *Anim. Front.* 9, 12–19. doi: 10.1093/af/vfy031

Cook, D. N., Kang, H. S., and Jetten, A. M. (2015). Retinoic acid-related orphan receptors (RORs): regulatory functions in immunity, development, circadian rhythm, and metabolism. *Nucl. Recept. Res.* 2, 1–24. doi: 10.11131/2015/101185

Deeb, N., and Cahaner, A. (1999). The effects of naked neck genotypes, ambient temperature, and feeding status and their interactions on body temperature and performance of broilers. *Poult. Sci.* 78, 1341–1346. doi: 10.1093/ps/78.10.1341

Degl'innocenti, D., Marzocchini, R., Malentacchi, F., Ramazzotti, M., Raugei, G., and Ramponi, G. (2004). *ACYP1* gene possesses two alternative splicing forms that induce apoptosis. *IUBMB Life* 56, 29–33. doi: 10.1080/15216540310001654349

Degl'innocenti, D., Ramazzotti, M., Sarchielli, E., Monti, D., Chevanne, M., Vannelli, G. B., et al. (2019). Oxadiazon affects the expression and activity of aldehyde dehydrogenase and acylphosphatase in human striatal precursor cells: a possible role in neurotoxicity. *Toxicology* 411, 110–121. doi: 10.1016/j.tox.2018.10.021

Felver-Gant, J., Mack, L., Dennis, R., Eicher, S., and Cheng, H. (2012). Genetic variations alter physiological responses following heat stress in 2 strains of laying hens. *Poult. Sci.* 91, 1542–1551. doi: 10.3382/ps.2011-01988

Fleming, D. S., Weigend, S., Simianer, H., Weigend, A., Rothschild, M., Schmidt, C., et al. (2017). Genomic comparison of indigenous African and Northern European chickens reveals putative mechanisms of stress tolerance related to environmental selection pressure. *G3: Genes Genomes Genet.* 7, 1525–1537. doi: 10.1534/g3.117.041228

Gorostizaga, A., Brion, L., Maloberti, P., Maciel, F. C., Podestá, E. J., and Paz, C. (2005). Heat shock triggers MAPK activation and MKP-1 induction in Leydig testicular cells. *Biochem. Biophys. Res. Commun.* 327, 23–28. doi: 10.1016/j bbrc.2004.11.129

Harding, H. P., Zhang, Y., and Ron, D. (1999). Protein translation and folding are coupled by an endoplasmic-reticulum–resident kinase. *Nature* 397, 271. doi: 10.1038/16729

Harris, S. L., and Levine, A. J. (2005). The p53 pathway: positive and negative feedback loops. *Oncogene* 24, 2899. doi: 10.1038/sj.onc.1208615

Hartl, F. (1996). Molecular chaperones in cellular protein folding. *Nature* 381, 571–579. doi: 10.1038/381571a0

Huang, C., Jiao, H., Song, Z., Zhao, J., Wang, X., and Lin, H. (2015). Heat stress impairs mitochondria functions and induces oxidative injury in broiler chickens. *J. Anim. Sci.* 93, 2144–2153. doi: 10.2527/jas.2014-8739

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44. doi: 10.1038/nprot.2008.211

Iwamoto, S., Koike, Y., Hosomichi, K., Hara, H., Yoshida, Y., Ogawa, H., et al. (2005). Identification of cDNA for the HSPA2, HSPA5 and HSPA8 orthologs of the heat shock protein 70 family from guinea fowl (Numida meleagris). *Anim. Sci. J.* 76, 519–524. doi: 10.1111/j.1740-0929.2005.00299.x

Janke, O., Tzschentke, B., and Boerjan, M. (2004). Comparative investigations of heat production and body temperature in embryos of modern chicken breeds. *Avian Poult. Biol. Rev.* 15, 191–196. doi: 10.3184/147020604783637868

Jastrebski, S. F., Lamont, S. J., and Schmidt, C. J. (2017). Chicken hepatic response to chronic heat stress using integrated transcriptome and metabolome analysis. *PloS One* 12, e0181900. doi: 10.1371/journal.pone.0181900

Juo, L.-Y., Liao, W.-C., Shih, Y.-L., Yang, B.-Y., Liu, A.-B., and Yan, Y.-T. (2016). HSPB7 interacts with dimerized FLNC and its absence results in progressive myopathy in skeletal muscles. *J. Cell Sci.* 129, 1661–1670. doi: 10.1242/jcs.179887

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kantanen, J., Løvendahl, P., Strandberg, E., Eythorsdottir, E., Li, M.-H., Kettunen-Præbel, A., et al. (2015). Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. *Front. Genet.* 6, 52. doi: 10.3389/fgene.2015.00052

Keck, M., Van Dijk, R. M., Deeg, C. A., Kistler, K., Walker, A., Von Rüden, E.-L., et al. (2018). Proteomic profiling of epileptogenesis in a rat model: focus on cell stress, extracellular matrix and angiogenesis. *Neurobiol. Dis.* 112, 119–135. doi: 10.1016/j.nbd.2018.01.013

Kilkenny, M. L., Doré, A. S., Roe, S. M., Nestoras, K., Ho, J. C., Watts, F. Z., et al. (2008). Structural and functional analysis of the Crb2–BRCT2 domain reveals distinct roles in checkpoint signaling and DNA damage repair. *Genes Dev.* 22, 2034–2047. doi: 10.1101/gad.472808

Kim, D., Langmead, B., and Salzberg, S. (2016). "HISAT2 implementation".

Kim, J.-S., Lee, Y.-H., Chang, Y.-U., and Yi, H.-K. (2017). PPARγ regulates inflammatory reaction by inhibiting the MAPK/NF-κB pathway in C2C12 skeletal muscle cells. *J. Physiol. Biochem.* 73, 49–57. doi: 10.1007/s13105-016-0523-3

Kingori, A., Wachira, A., and Tuitoek, J. (2010). Indigenous chicken production in Kenya: a review. *Int. J. Poult. Sci.* 9, 309–316. doi: 10.3923/ijps.2010.309.316

Kurebayashi, S., Ueda, E., Sakaue, M., Patel, D. D., Medvedev, A., Zhang, F., et al. (2000). Retinoid-related orphan receptor γ (RORγ) is essential for lymphoid organogenesis and controls apoptosis during thymopoiesis. *Proc. Natl. Acad. Sci.* 97, 10132–10137. doi: 10.1073/pnas.97.18.10132

Lan, X., Hsieh, J. C., Schmidt, C. J., Zhu, Q., and Lamont, S. J. (2016). Liver transcriptome response to hyperthermic stress in three distinct chicken lines. *BMC Genomics* 17, 955. doi: 10.1186/s12864-016-3291-0

Lara, L., and Rostagno, M. (2013). Impact of heat stress on poultry production. *Animals* 3, 356–369. doi: 10.3390/ani3020356

Lawrence, A., and Wall, E. (2014). Selection for 'environmental fit' from existing domesticated species. *Rev. Sci. Tech.* 33, 171–179. doi: 10.20506/rst.33.1.2275

Liu, L., Fu, C., Yan, M., Xie, H., Li, S., Yu, Q., et al. (2016). Resveratrol modulates intestinal morphology and HSP70/90, NF-κB and EGF expression in the jejunal mucosa of black-boned chickens on exposure to circular heat stress. *Food Funct.* 7, 1329–1338. doi: 10.1039/C5FO01338K

Loyau, T., Bedrani, L., Berri, C., Métayer-Coustard, S., Praud, C., Coustham, V., et al. (2015). Cyclic variations in incubation conditions induce adaptive responses to later heat exposure in chickens: a review. *Animal* 9, 76–85. doi: 10.1017/S1751731114001931

Loyau, T., Hennequet-Antier, C., Coustham, V., Berri, C., Leduc, M., Crochet, S., et al. (2016). Thermal manipulation of the chicken embryo triggers differential gene expression in response to a later heat challenge. *BMC Genomics* 17, 329. doi: 10.1186/s12864-016-2661-y

Loyau, T., Métayer-Coustard, S., Berri, C., Crochet, S., Cailleau-Audouin, E., Sannier, M., et al. (2014). Thermal manipulation during embryogenesis has long-term effects on muscle and liver metabolism in fast-growing chickens. *PLoS One* 9, e105339. doi: 10.1371/journal.pone.0105339

Lu, Q., Wen, J., and Zhang, H. (2007). Effect of chronic heat exposure on fat deposition and meat quality in two genetic types of chicken. *Poult. Sci.* 86, 1059–1064. doi: 10.1093/ps/86.6.1059

Magothe, T., Okeno, T., Muhuyi, W., and Kahi, A. (2012). Indigenous chicken production in Kenya: I. Current status. *World's Poult. Sci. J.* 68, 119–132. doi: 10.1017/S0043933912000128

Mcmichael, A. J., Powles, J. W., Butler, C. D., and Uauy, R. (2007). Food, livestock production, energy, climate change, and health. *Lancet* 370, 1253–1263. doi: 10.1016/S0140-6736(07)61256-2

Mekonnen, H., Mulatu, D., Kelay, B., and Berhan, T. (2010). Assessment of the nutritional status of indigenous scavenging chickens in Ada'a district, Ethiopia. *Trop. Anim. Health Prod.* 42, 123. doi: 10.1007/s11250-009-9395-7

Monson, M. S., Van Goor, A. G., Ashwell, C. M., Persia, M. E., Rothschild, M. F., Schmidt, C. J., et al. (2018). Immunomodulatory effects of heat stress and lipopolysaccharide on the bursal transcriptome in two distinct chicken lines. *BMC Genomics* 19, 643. doi: 10.1186/s12864-018-5033-y

Moraa, G., Oyier, P., Maina, S., Makanda, M., Ndiema, E., Alakonya, A., et al. (2015). Assessment of phenotypic traits relevant for adaptation to hot environments in indigenous chickens from four agro-climatic zones of Kenya. *Livest. Res. Rural Dev.* 27, 1–11.

Muiruri, H., and Harrison, P. (1991). Effect of roost temperature on performance of chickens in hot ambient environments. *Poult. Sci.* 70, 2253–2258. doi: 10.3382/ps.0702253

Mujahid, A., Pumford, N. R., Bottje, W., Nakagawa, K., Miyazawa, T., Akiba, Y., et al. (2007). Mitochondrial oxidative damage in chicken skeletal muscle induced by acute heat stress. *J. Poult. Sci.* 44, 439–445. doi: 10.2141/jpsa.44.439

Mujahid, A., Sato, K., Akiba, Y., and Toyomizu, M. (2006). Acute heat stress stimulates mitochondrial superoxide production in broiler skeletal muscle, possibly *via* downregulation of uncoupling protein content. *Poult. Sci.* 85, 1259–1265. doi: 10.1093/ps/85.7.1259

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., and Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livest. Sci.* 130, 57–69. doi: 10.1016/j.livsci.2010.02.011

Njarui, D., Gichangi, E., Gatheru, M., Nyambati, E., Ondiko, C., Njunie, M., et al. (2016). A comparative analysis of livestock farming in smallholder mixed crop-livestock systems in Kenya: 1. Livestock inventory and management. *Development* 28, 4.

Nyaga, P. (2007). Good biosecurity practices in small scale commercial and scavenging production systems in Kenya. *Rome: FAO* 157, 1–34.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29

Ouko, V., Ilatsia, E., Oduho, G., and Kios, D. (2017). Genetic parameters for large white pigs reared under intensive management systems in Kenya. *East Afr. Agric. For. J.* 82, 47–56. doi: 10.1080/00128325.2016.1219544

Padgett, D. A., and Glaser, R. (2003). How stress influences the immune response. *Trends Immunol.* 24, 444–448. doi: 10.1016/S1471-4906(03)00173-X

Park, W., Srikanth, K., Lim, D., Park, M., Hur, T., Kemp, S., et al. (2019). Comparative transcriptome analysis of Ethiopian indigenous chickens from low and high altitudes under heat stress condition reveals differential immune response. *Anim. Genet.* 50, 42–53. doi: 10.1111/age.12740

Pearson, G., Robinson, F., Beers Gibson, T., Xu, B.-E., Karandikar, M., Berman, K., et al. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.* 22, 153–183. doi: 10.1210/edrv.22.2.0428

Porto-Neto, L. R., Reverter, A., Prayaga, K. C., Chan, E. K., Johnston, D. J., Hawken, R. J., et al. (2014). The genetic architecture of climatic adaptation of tropical cattle. *PloS One* 9, e113284. doi: 10.1371/journal.pone.0113284

Rappold, I., Iwabuchi, K., Date, T., and Chen, J. (2001). Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage–signaling pathways. *J. Cell Biol.* 153, 613–620. doi: 10.1083/jcb.153.3.613

Renaudeau, D., Collin, A., Yahav, S., De Basilio, V., Gourdine, J., and Collier, R. (2012). Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Animal* 6, 707–728. doi: 10.1017/S1751731111002448

Reverter, A., and Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. doi: 10.1093/bioinformatics/btn482

Rothschild, M. F., and Plastow, G. S. (2014). Applications of genomics to improve livestock in the developing world. *Livest. Sci.* 166, 76–83. doi: 10.1016/j.livsci.2014.03.020

Sandercock, D., Hunter, R., Nute, G., Mitchell, M., and Hocking, P. (2001). Acute heat stress–induced alterations in blood acid–base status and skeletal muscle membrane integrity in broiler chickens at two ages: implications for meat quality. *Poult. Sci.* 80, 418–425. doi: 10.1093/ps/80.4.418

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C T method. *Nat. Protoc.* 3, 1101. doi: 10.1038/nprot.2008.73

Seebacher, F. (2009). Responses to temperature variation: integration of thermoregulation and metabolism in vertebrates. *J. Exp. Biol.* 212, 2885–2891. doi: 10.1242/jeb.024430

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Silvestri, S., Bryan, E., Ringler, C., Herrero, M., and Okoba, B. (2012). Climate change perception and adaptation of agro-pastoral communities in Kenya. *Reg. Environ. Change* 12, 791–802. doi: 10.1007/s10113-012-0293-6

Soleimani, A., and Zulkifli, I. (2010). Effects of high ambient temperature on blood parameters in Red Jungle fowl, Village fowl and broiler chickens. *J. Anim. Vet. Adv.* 9, 1201–1207. doi: 10.3923/javaa.2010.1201.1207

Sonaiya, E. (1990). The context and procespects for development of smallholder rural poultry production in Africa. *Proceedings of CTA-Seminar on Smallholder Rural Poultry Production*; October 9–13; Thessaloniki, Greece 1, 35–52.

Srikanth, K., Kwon, A., Lee, E., and Chung, H. (2017a). Characterization of genes and pathways that respond to heat stress in Holstein calves through transcriptome analysis. *Cell Stress Chaperones* 22, 29–42. doi: 10.1007/s12192-016-0739-8

Srikanth, K., Lee, E., Kwan, A., Lim, Y., Lee, J., Jang, G., et al. (2017b). Transcriptome analysis and identification of significantly differentially expressed genes in Holstein calves subjected to severe thermal stress. *Int. J. Biometeorol.* 61, 1993–2008. doi: 10.1007/s00484-017-1392-3

Sun, H., Jiang, R., Xu, S., Zhang, Z., Xu, G., Zheng, J., et al. (2015a). Transcriptome responses to heat stress in hypothalamus of a meat-type chicken. *J. Anim. Sci. Biotechnol.* 6, 6. doi: 10.1186/s40104-015-0003-6

Sun, L., Lamont, S. J., Cooksey, A. M., Mccarthy, F., Tudor, C. O., Vijay-Shanker, K., et al. (2015b). Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line. *Cell Stress Chaperones* 20, 939–950. doi: 10.1007/s12192-015-0621-0

Te Pas, M. F., Park, W., Srikanth, K., Kemp, S., Kim, J.-M., Lim, D., et al. (2019). Transcriptomic profiles of muscle, heart, and spleen in reaction to circadian heat stress in Ethiopian highland and lowland male chicken. *Cell Stress Chaperones* 24, 175–194. doi: 10.1007/s12192-018-0954-6

Team, R. C. (2013). R: a language and environment for statistical computing.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562. doi: 10.1038/nprot.2012.016

Vallanat, B., Anderson, S. P., Brown-Borg, H. M., Ren, H., Kersten, S., Jonnalagadda, S., et al. (2010). Analysis of the heat shock response in mouse liver reveals transcriptional dependence on the nuclear receptor peroxisome proliferator-activated receptor α (PPARα). *BMC Genomics* 11, 16. doi: 10.1186/1471-2164-11-16

Vogelstein, B., Lane, D., and Levine, A. J. (2000). Surfing the p53 network. *Nature* 408, 307. doi: 10.1038/35042675

Vos, M. J., Zijlstra, M. P., Kanon, B., Van Waarde-Verhagen, M. A., Brunt, E. R., Oosterveld-Hut, H. M., et al. (2010). HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Hum. Mol. Genet.* 19, 4677–4693. doi: 10.1093/hmg/ddq398

Wang, M., An, S., Wang, D., Ji, H., Geng, M., Guo, X., et al. (2018). Quantitative proteomics identify the possible tumor suppressive role of protease-activated receptor-4 in esophageal squamous cell carcinoma cells. *Pathol. Oncol. Res.*, 3, 1–7. doi: 10.1007/s12253-018-0395-7

Wang, S.-H., Cheng, C.-Y., Tang, P.-C., Chen, C.-F., Chen, H.-H., Lee, Y.-P., et al. (2015). Acute heat stress induces differential gene expressions in the testes of a broiler-type strain of Taiwan country chickens. *PLoS One* 10, e0125816. doi: 10.1371/journal.pone.0125816

Wang, Y.-X. (2010). PPARs: diverse regulators in energy metabolism and metabolic diseases. *Cell Res.* 20, 124. doi: 10.1038/cr.2010.13

Wang, Y., Saelao, P., Chanthavixay, K., Gallardo, R., Bunn, D., Lamont, S., et al. (2017). Physiological responses to heat stress in two genetically distinct chicken inbred lines. *Poult. Sci.* 97, 770–780. doi: 10.3382/ps/pex363

Welihinda, A. A., Tirasophon, W., and Kaufman, R. J. (1999). The cellular response to protein misfolding in the endoplasmic reticulum. *Gene Expr.* 7, 293–300.

Wenzl, K., Troppan, K., Neumeister, P., and Ja Deutsch, A. (2015). The nuclear orphan receptor NR4A1 and NR4A3 as tumor suppressors in hematologic neoplasms. *Curr. Drug Targets* 16, 38–46. doi: 10.2174/1389450115666141120112818

Wolfenson, D., Bachrach, D., Maman, M., Graber, Y., and Rozenboim, I. (2001). Evaporative cooling of ventral regions of the skin in heat-stressed laying hens. *Poult. Sci.* 80, 958–964. doi: 10.1093/ps/80.7.958

Xia, G., Evers, M. R., Kang, H.-G., Schachner, M., and Baenziger, J. U. (2000). Glycobiology and extracellular matrices—molecular cloning and expression of the pituitary glycoprotein hormone N-acetylgalactosamine-4-O-sulfotransferase. *J. Biol. Chem.* 275, 38402–38409. doi: 10.1074/jbc.M007821200

Xie, J., Tang, L., Lu, L., Zhang, L., Lin, X., Liu, H.-C., et al. (2015). Effects of acute and chronic heat stress on plasma metabolites, hormones and oxidant status in restrictedly fed broiler breeders. *Poult. Sci.* 94, 1635–1644. doi: 10.3382/ps/pev105

Yang, Y.-L., Loh, K.-S., Liou, B.-Y., Chu, I.-H., Kuo, C.-J., Chen, H.-D., et al. (2013). SESN-1 is a positive regulator of lifespan in *Caenorhabditis elegans*. *Exp. Gerontol.* 48, 371–379. doi: 10.1016/j.exger.2012.12.011

Yunis, R., and Cahaner, A. (1999). The effects of the naked neck (Na) and frizzle (F) genes on growth and meat yield of broilers and their interactions with ambient temperatures and potential growth rate. *Poult. Sci.* 78, 1347–1352. doi: 10.1093/ps/78.10.1347

Zhang, J., Schmidt, C. J., and Lamont, S. J. (2017). Transcriptome analysis reveals potential mechanisms underlying differential heart development in fast-and slow-growing broilers under heat stress. *BMC Genomics* 18, 295. doi: 10.1186/s12864-017-3675-9

Zhang, X.-D., Qin, Z.-H., and Wang, J. (2010). The role of p53 in cell metabolism. *Acta Pharmacol. Sin.* 31, 1208. doi: 10.1038/aps.2010.151

Zhao, X., Wu, Q., Fu, X., Yu, B., Shao, Y., Yang, H., et al. (2010). Examination of copy number variations of CHST9 in multiple types of hematologic malignancies. *Cancer Genet. Cytogenet.* 203, 176–179. doi: 10.1016/j.cancergencyto.2010.07.132

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Srikanth, Kumar, Park, Byun, Lim, Kemp, te Pas, Kim and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

CORRECTION
published: 03 March 2020
doi: 10.3389/fgene.2020.00197

Corrigendum: Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response

*Krishnamoorthy Srikanth[1], Himansu Kumar[1], Woncheoul Park[1], Mijeong Byun[1], Dajeong Lim[1], Steve Kemp[2], Marinus F. W. te Pas[3], Jun-Mo Kim[4] and Jong-Eun Park[1]**

[1] Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Wanju, South Korea, [2] Animal Biosciences, International Livestock Research Institute (ILRI), Nairobi, Kenya, [3] Wageningen UR Livestock Research, Animal Breeding and Genomics, Wageningen, Netherlands, [4] Department of Animal Science and Technology, Chung-Ang University, Anseong, South Korea

Keywords: heat stress, hub genes, PPAR signaling, MAPK signaling, p53 signaling, RNA-Seq

OPEN ACCESS

Edited and reviewed by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

****Correspondence:***
Jong-Eun Park
jepark0105@korea.kr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *09 December 2019*
Accepted: *19 February 2020*
Published: *03 March 2020*

Citation:
Srikanth K, Kumar H, Park W, Byun M, Lim D, Kemp S, te Pas MFW, Kim J-M and Park J-E (2020) Corrigendum: Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response. Front. Genet. 11:197. doi: 10.3389/fgene.2020.00197

A Corrigendum on

Cardiac and Skeletal Muscle Transcriptome Response to Heat Stress in Kenyan Chicken Ecotypes Adapted to Low and High Altitudes Reveal Differences in Thermal Tolerance and Stress Response

by Srikanth, K., Kumar, H., Park, W., Byun, M., Lim, D., Kemp, S., et al. (2019). Front. Genet. 10:993. doi: 10.3389/fgene.2019.00993

In the original article, there was a mistake in **Supplementary Table 1** and **Supplementary Table 2**. The expression values given for *PDK4* in Supplementary Table 1, ALL_M, CLL_M, AHL_M, CHL_M contrasts were −3.90808, 2.10011, −4.12057, and −4.12057 the correct values are −4.1009, 2.07292, −4.63904, and 3.05659 same value should appear at the "Max_expression_level" column in HL_node_table in Supplementary Table 2.

Similarly the expression values of *MT4* given in Supplementary Table 1 for ALL_H, CLL_H, AHL_H, and CHL_H are −1.20147, 1.485881, −1.19557, and 1.0025, the correct values are −3.1675, −1.82983, −1.35669, and −1.84142. To reflect this change, columns "Max_expression_level," "Max_Tissue," and "Up/Down" on Supplementary Table 2, "LL_node_table" tab is corrected.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Copyright © 2020 Srikanth, Kumar, Park, Byun, Lim, Kemp, te Pas, Kim and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 October 2019
doi: 10.3389/fgene.2019.01040

Identification of lncRNAs by RNA Sequencing Analysis During *in Vivo* Pre-Implantation Developmental Transformation in the Goat

Ying-hui Ling[1,2†], Qi Zheng[1,2†], Yun-sheng Li[1,2†], Meng-hua Sui[1,2], Hao Wu[1,2], Yun-hai Zhang[1,2*], Ming-xing Chu[3], Yue-hui Ma[3], Fu-gui Fang[1,2] and Li-na Xu[1,4]*

[1] College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [2] Local Animal Genetic Resources Conservation and Biobreeding Laboratory of Anhui Province, Hefei, China, [3] Key Laboratory of Farm Animal Genetic Resources and Germplasm Innovation of Ministry of Agriculture, Chinese Academy of Agricultural Sciences, Beijing, China, [4] Institute of Plant Protection and Agro-Products Safety, Anhui Academy of Agricultural Sciences, Hefei, China

Pre-implantation development is a dynamic, complex and precisely regulated process that is critical for mammalian development. There is currently no description of the role of the long noncoding RNAs (lncRNAs) during the pre-implantation stages in the goat. The *in vivo* transcriptomes of oocytes (n = 3) and pre-implantation stages (n=19) at seven developmental stages in the goat were analyzed by RNA sequencing (RNA-Seq). The major zygotic gene activation (ZGA) event was found to occur between the 8- and 16-cell stages in the pre-implantation stages. We identified 5,160 differentially expressed lncRNAs (DELs) in developmental stage comparisons and functional analyses of the major and minor ZGAs. Fourteen lncRNA modules were found corresponding to specific pre-implantation developmental stages by weighted gene co-expression network analysis (WGCNA). A comprehensive analysis of the lncRNAs at each developmental transition of high correlation modules was done. We also identified lncRNA-mRNA networks and hub-lncRNAs for the high correlation modules at each stage. The extensive association of lncRNA target genes with other embryonic genes suggests an important regulatory role for lncRNAs in embryonic development. These data will facilitate further exploration of the role of lncRNAs in the developmental transformation in the pre- implantation stage.

Keywords: RNA-seq, Goat, long noncoding RNAs, pre-implantation development, zygotic gene activation

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Ravi Kumar Gandham, Indian Veterinary Research Institute (IVRI), India
Stephen J. Bush, University of Oxford, United Kingdom
Hongping Zhang, Sichuan Agricultural University, China,
Feng Wang, Nanjing Agricultural University, China

***Correspondence:**
Ying-hui Ling
lingyinghui@ahau.edu.cn
Yun-hai Zhang
zhangyunhai01@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *19 May 2019*
Accepted: *30 September 2019*
Published: *25 October 2019*

Citation:
Ling Y-h, Zheng Q, Li Y-s, Sui M-h, Wu H, Zhang Y-h, Chu M-x, Ma Y-h, Fang F-g and Xu L-n (2019) Identification of lncRNAs by RNA Sequencing Analysis During in Vivo Pre-Implantation Developmental Transformation in the Goat. Front. Genet. 10:1040. doi: 10.3389/fgene.2019.01040

INTRODUCTION

Pre-implantation development comprises complex and dynamic regulatory processes involving specific and stable gene expression patterns that maintain the viability of the embryo. During different embryonic stages, highly complex tissues are composed of different cell types that are formed by cell fate and cell differentiation (Lokken and Ralston, 2016; Bissiere et al., 2018). Analysis of the spatiotemporal patterns of gene expression in goat pre-implantation stages is therefore essential for clarifying early developmental processes in this species. The key stage in the transition from germ cells to embryonic development is zygotic gene activation (ZGA), which induces developmental blocks of embryonic development (Lee et al., 2014). The timing of mammalian ZGA process is species-specific; it occurs from the point of oocyte maturation until mRNA transcriptional activity in the embryo. The initiation of major ZGA events have been reported to be the 2-cell stage in mouse (Xue et al., 2013), 4-cell stage in pig (Cao et al., 2014), and

4-cell to 8-cell stage in human (Xue et al., 2013; Yan et al., 2013). Previous studies had reported that ZGA-related genes begin to be expressed in the 8-cell to 16-cell stage of goats (Ma et al., 2014; Deng et al., 2018). Recent study of the developmental block of cultured of goat *in vitro* suggested that cell development stops at the 8-cell stage, and further verification by RNA-seq has indicated that it occurs between the 4- and 8-cell stage (Deng et al., 2018).

Encoded protein sequences represent less than 2% of the mammalian genome whereas a much larger fraction of this genome is transcribed into what is known as noncoding RNAs (ncRNAs) (Agliano et al., 2019). Many ncRNAs are expressed in pre-implantation stages and play an important role in fertilization and appropriate embryonic development (Hamazaki et al., 2015; Yuan et al., 2016; Vallot et al., 2017). Long ncRNAs (lncRNAs) are among the largest ncRNAs in vertebrates and are broadly defined as noncoding transcripts of greater than 200 nucleotides (Sun and Kraus, 2013; Agliano et al., 2019). For example, Trincr1 binds to TRIM71 to inhibit FGF/ERK signaling in embryonic stem cells to coordinate cell fate specifications (Li et al., 2019). Most of the studies on the expression of lncRNAs in pre-implantation stages have been focused on humans (Kurian et al., 2015), mice (Karlic et al., 2017), and pigs (Zhong et al., 2018). In comparable functional studies of oocyte and pre-implantation cells lncRNAs in the goat are limited.

The domestic goat (*Capra hircus*) is one of the most important commercially farmed animals that produces a variety of products, including meat, milk, and skins (Guan et al., 2016). Moreover, various established reproductive biotechnologies have made the goat a significant species used in agriculture and transgenic breeding research (Baguisi et al., 1999; Bao et al., 2016). The emergence of low input high throughput sequencing technologies has enabled the transcriptome to be determined from oocytes and pre-implantation cells at different stages of development in the goat.

In our current study, the transcriptomes of seven pre-implantation developmental stages of goat, including *in vivo* metaphase II mature oocytes and the 2-cell, 4-cell, 8-cell, 16-cell, morula and blastocyst stages, were sequenced using low input high throughput RNA-seq. This analysis identified the timing of goat ZGA and identified the differential expression of lncRNAs in oocytes and pre-implantation stages, and thereby revealed the role of the lncRNAs in ZGA event. Further, we constructed a WGCNA network to identify the lncRNAs and lncRNA-mRNA networks that are highly correlated at each stage, and to identify the hub-lncRNAs in all pre-implantation stages. This compilation-specific network analysis has given us a more comprehensive understanding of the functional transition of lncRNAs at specific stages of pre-implantation in the goat.

MATERIALS AND METHODS

Goat Pre-Implantation Stages Material

Female Anhui white goats (AWGs) were farm-raised by the Boda Company (Baogong Town, Feidong County, Hefei, China) under a unified field management system. All experimental animals were estrus-synchronized by treatment with EAZI-Breed CIDR (CIDR, Hamilton, New Zealand) for 12 days and superovulated prior to CIDR removal. The estrus test was performed 12h after stopping CIDR, and artificial insemination was performed on the female AWGs that were estrus at the same time. After 36–48, 56–60, 87–92, 97–100, and 109–112 h of mating, oocytes and 2-cell, 4-cell, 8-cell, and 16-cell cells were flushed from the oviduct. Morulae and blastocysts were obtained from the uterus after 152–156 and 212–218 h, respectively. A total of 21 samples were obtained in these seven stages, and each stage of the sample had three replicates. Oocytes and pre-implantation cells were washed several times in 1% DBPS solution. Five obtained oocytes and pre-implantation cells at each stage were pooled and snap frozen in liquid nitrogen.

RNA Isolation, Library Preparation, and Sequencing

RNA isolation, library construction and sequencing were done by Novogene Co. Ltd. (Beijing, China). Total RNA from individual oocytes and pre-implantation cells was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA); and RNA was co-precipitated with linear acrylamide (Ambion, Texas, USA). RNA integrity was evaluated on 1% agarose gel. RNA purity was checked using a NanoPhotometer (Implen, CA, USA). RNA concentrations were measured using a Qubit® RNA Assay Kit and Qubit® 2.0 Flurometer (Life Technologies, CA, USA). We then used 3 ng of RNA as the base material for cDNA sample preparation, and purified cDNA was obtained and detected on an Agilent Bioanalyzer 2100 system (Agilent technologies, CA, USA). The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia, CA, USA) in accordance with the manufacturer's instructions. After cluster generation, the libraries were sequenced on an Illumina Hiseq 2500 platform and 150 bp paired-end reads were generated (**Table S1**).

Data Analysis

Raw data (raw reads) in a fastq format were first processed through in-house perl scripts (ng-qc). All the linker sequences in the raw data would be removed, ng-qc parameter: -L 20 -p 0.5 (-L, lowest quality value, -p parameter of low-quality reads.-L20 –p 0.5 was the low-quality base ratio allowed by the specified reads; the default was 0.5. This means that the number of bases of quality value ≤ -L parameter (Baguisi et al., 1999)/ reads length ≥ 0.5 represented low quality reads). In addition, entering the adapter sequence in the ng-qc software would be removed by sequence matching. Clean data (clean reads) were obtained by removing reads from the raw data that contained adapters, reads with undetermined base content greater than 10%, and low-quality reads (**Table S1**). Moreover, clean reads satisfied the conditions of Q20 > 90% and Q30 > 85%. This meant that reads with a base error rate of less than 0.01 account for more than 90% of all reads, and reads with an error rate of less than 0.001 account for more than 85% of all reads. The *Capra hircus* reference genome and gene model annotation files

Abbreviations: ZGA, zygotic gene activation; DEL, differentially expressed lncRNA; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; WGCNA, Weighted gene co-expression network analysis; TOM, topological overlap matrix.

for this study can be accessed at (The Capra hircus reference gene model annotation file; The Capra hircus reference genome model annotation file). An index of the reference genome was built using Bowtie v2.0.6 (Langmead and Salzberg, 2012) and paired-end clean reads were aligned to this using TopHat v2.0.9, both with default parameters (Trapnell et al., 2009). The mapped reads of each sample were assembled using both Scripture (beta2) (Trapnell et al., 2010) and Cufflinks (v2.1.1) (Guttman et al., 2010) *via* a reference-based approach. Scripture used a statistical segmentation model to distinguish expressed loci from experimental noise and spliced reads to assemble expressed segments. It reported all statistically expressed isoforms in a given locus. Cufflinks uses a probabilistic model to simultaneously assemble and quantify the expression level of a minimal set of isoforms that provides a maximum likelihood explanation of the expression data in a given locus. Scripture was run with default parameters, Cufflinks was run with 'min-frags-per-transfrag = 0' and '–library-type'; other parameters were set as default.

Based on the splicing results, the structural characteristics of lncRNA and the functional characteristics of non-encoded proteins, a 5-step screening was performed, and the lncRNAs obtained were used as the final candidate lncRNA set for subsequent analysis. First, the transcripts spliced from all samples were combined using cuffcompare to screen for transcripts of unknown molecular orientation. Second, we chose transcripts with transcript length ≥ 200 bp and exon number ≥ 2. Then, we calculated the read coverage of each transcript by cufflinks and selected a transcript with a coverage of ≥ 3 reads in at least one sample. Next, the transcript obtained in the previous step was first compared with the known lncRNA by cuffcompare to obtain the same transcript as the known lncRNA. This part of the transcript was directly included in the final lncRNA set and no further screening was performed. Finally, the transcripts of the candidate lincRNA, intronic lncRNA, and anti-sense lncRNA type were screened by comparison with known mRNAs and using the class_code information in the cuffcompare analysis results (**Table S2**) (Cuffcompare program of Cufflinks).

Then, transcripts with coding potential were filtered by Coding-Non-Coding-Index (CNCI) (v2) (Sun et al., 2013), Coding Potential Calculator (CPC) (0.9-r2) (Kong et al., 2007), and Pfam-scan (PFAM) (v1.3) (31), and the noncoding transcripts were selected as our candidate lncRNAs. The CNCI parameters include –f input transcriptome sequence file, –o data output path, –p 1 (number of cpu) and -m ve (specified mode, ve is vertebrate). The index in the CNCI prediction result would be labeled as coding or noncoding (Sun et al., 2013). CPC (0.9-r2), used with default parameters, searched sequences with known protein sequence databases to elucidate both coding and non-coding transcripts (Kong et al., 2007). In addition, we translated each transcript in all three possible frames and used PFAM (v1.3) to identify the presence of any known protein family domain recorded in the Pfam database (release 27; used Pfam B). A transcript with a PFAM hit will be excluded in the following steps. Pfam searches use default parameters of -E 0.001 -domE 0.001 and -cpu 2 (CPU set to 2) (Bateman et al., 2004; Punta et al., 2012).

Quantification of Gene Expression and Differential Expression Analysis

Cuffdiff (v2.1.1) was used to calculate the FPKM (fragments per kilo-base of exon per million fragments mapped) of both the lncRNAs and coding genes in each sample (Trapnell et al., 2010). Gene FPKMs were computed by summing those for the transcripts in each gene. Principal Component Analysis (PCA) was conducted using R and heat map/cluster analysis using the Morpheus free online platform (Morpheus). The applied statistical procedures used a negative binomial distribution model in Cuffdiff to determine differentially expressed transcripts (Trapnell et al., 2010). For biological replicates, transcripts or genes with a P-adj < 0.05 were assigned as differentially expressed.

Target Gene Prediction and Functional Analysis

The interaction of lncRNA with a nearby target gene was called *cis*- action. We searched for coding genes 10 kb upstream and downstream of each lncRNA. Candidate target genes for *trans*-acting lncRNAs were predicted based on co-expression. The Pearson correlation coefficient method was used to analyze correlations between mRNAs and lncRNAs. mRNAs with absolute correlation value greater than 0.95 were considered to be target genes for lncRNAs. LncRNA-mRNA networks were constructed using Cytospace (Cytoscpace software). Gene Ontology (GO) is a classification system for internationally standardized gene functions that provides a controlled vocabulary to comprehensively describe the properties of genes and their products. GO enrichment analysis of differentially expressed genes or lncRNA target genes was performed using the GO-seq R package, in which gene length bias was corrected (Young et al., 2010). GO terms with P-value < 0.05 were considered to indicate significant enrichment of those respective differential genes. Bubble charts were constructed using the OmicShare platform for data analysis (Omicshare tools).

Weighted Gene Co-Expression Network Analysis (WGCNA)

Differentially expressed lncRNAs with an FPKM > 0.01 between all pre-implantation cells development stages were selected, and the lncRNA co-expression network was then constructed using R package WGCNA (Langfelder and Horvath, 2008). A signed weighted correlation network was generated by first creating a matrix of Pearson correlation coefficients between all pairs of genes across the measured samples. An adjacency matrix was then transformed into a topological overlap matrix (TOM) to minimize the effects of noise and spurious associations. To define modules as branches, we employed the Dynamic Tree Cut algorithm with default parameters to cut the hierarchal clustering tree (Langfelder et al., 2008).

Quantitative RT-PCR

QRT-PCR was performed using GoTaq qPCR Master Mix (Promega, Madison, WI) and Real-time Thermal Cycler 5100 (Thermo, Shanghai, China). The primer pairs used in the PCR

amplifications were synthesized by the Beijing Genomics Institute and are listed in **Table S3**. The GAPDH housekeeping gene was amplified as a control (Li et al., 2019). The target sequence levels were normalized to the reference sequence and calculated as $2^{-\Delta\Delta Ct}$. Statistical analysis of the normalized data was then conducted using SPSS version 19.0 for Windows (SPSS Statistics). Data are presented as means ± SEM. Data were considered statistically significant at P-value < 0.05.

RESULTS

Transcriptome Reconstruction From RNA-Seq Data

We collected 21 samples from Anhui white goats after superovulation treatment and then performed RNA-seq analysis (**Figure 1A**). The cells were obtained from seven crucial stages i.e. metaphase II oocytes and 2-cell, 4-cell, 8-cell, 16-cell, morula and blastocyst stage (**Figure 1B**). An Illumina HiSeq 2500 sequencer was used and 290.8 GB of clean sequencing data were generated from the 21 samples, with an average of 92.1 million total mapped reads per stage (**Table S1**).

Dynamic Patterns of Protein-Coding Transcript Profiles

A total of 29,608 protein-coding transcripts were identified during the seven goat pre-implantation stages (**Table S4**). Principal component analysis was used to capture the expression of transcripts from the oocyte to blastocyst development stages. Oocytes and pre-implantation cells at the same stage were found to cluster with each other, except that one 4-cell stage was clustered in the 2-cell stage, and one morula was clustered in the 16-cell stage (**Figure 2A**). The greatest changes in gene expression were observed in the 8- and 16-cell stages, possibly due to maternal-zygote transitions during this period. Hierarchical clustering also yielded similar intra- and inter-phase expression patterns (**Figure 2B**). All of the stages of goat development were divided into two

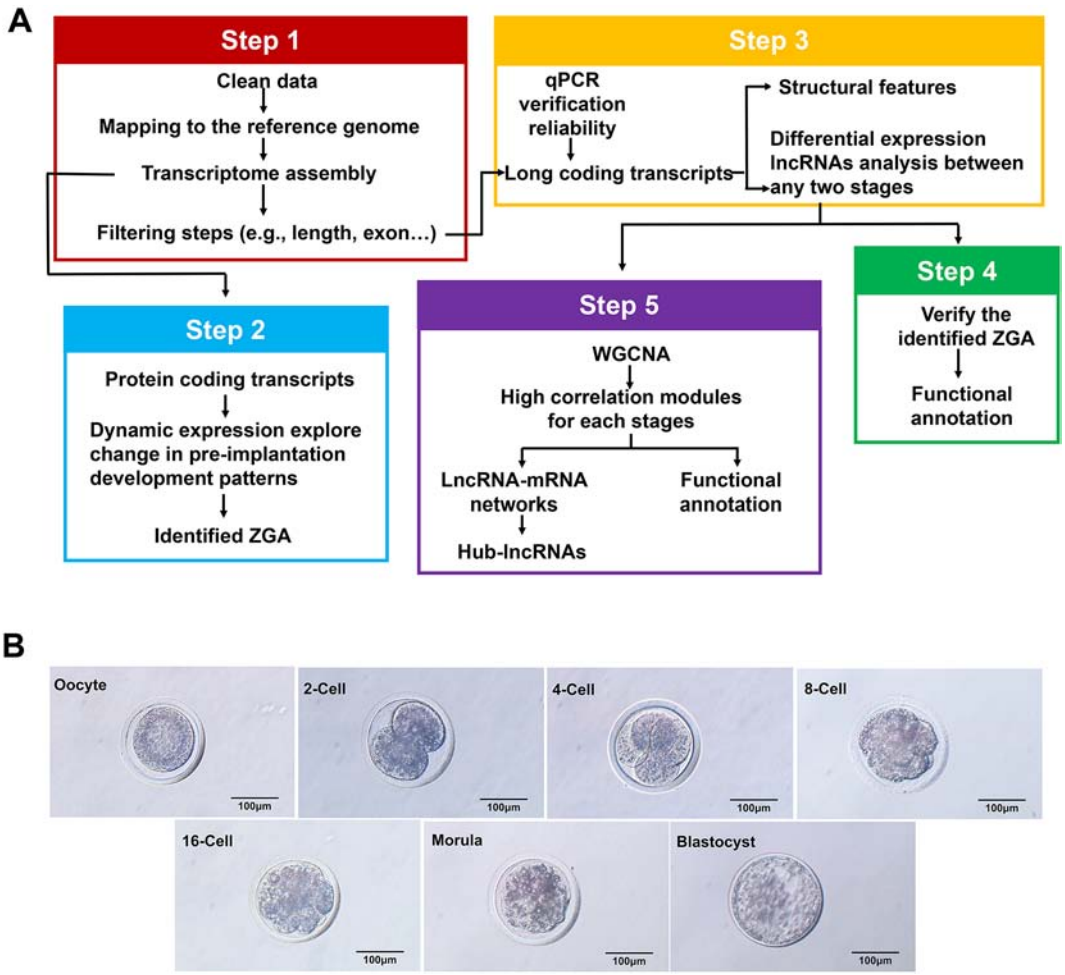
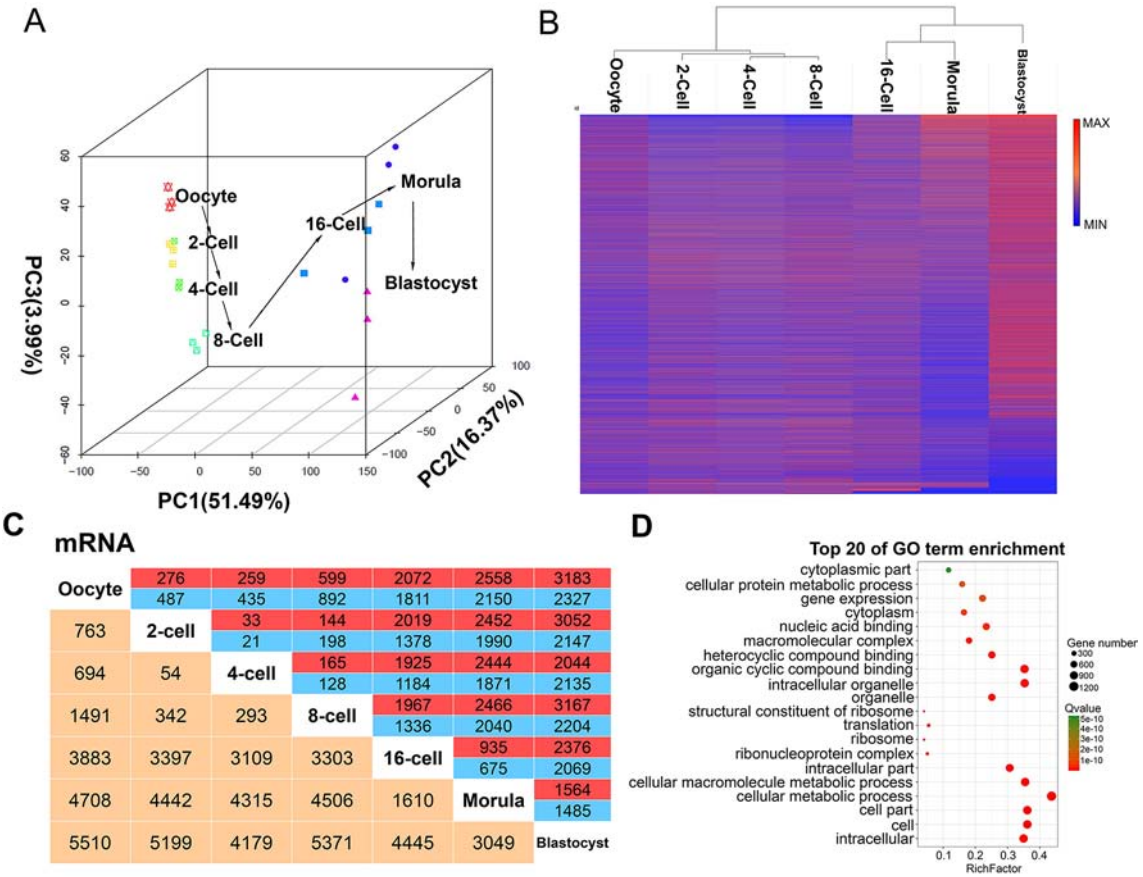


FIGURE 1 | Comprehensive catalog of the lncRNAs in goat oocytes and embryos at different stages. **(A)** Schematic of the experimental design method for the identification of expressed protein-coding transcripts and lncRNAs. Scale bar, 100 μm. **(B)** Microscopy imaging of mature goat oocytes and embryos at the metaphase II oocyte, 2-cell, 4-cell, 8-cell, 16-cell, morula and late blastocyst stages.



Oocyte	276	259	599	2072	2558	3183
	487	435	892	1811	2150	2327
763	2-cell	33	144	2019	2452	3052
		21	198	1378	1990	2147
694	54	4-cell	165	1925	2444	2044
			128	1184	1871	2135
1491	342	293	8-cell	1967	2466	3167
				1336	2040	2204
3883	3397	3109	3303	16-cell	935	2376
					675	2069
4708	4442	4315	4506	1610	Morula	1564
						1485
5510	5199	4179	5371	4445	3049	Blastocyst

FIGURE 2 | mRNA expression patterns during pre-implantation development in the goat. **(A)** Principal component analysis (PCA) of mRNAs in 21 goat pre-implantation development samples at 7 different stages. The same color represents the same stage. The arrows indicate the direction of development between successive muscle stages. **(B)** Hierarchical clustering heat map of mRNAs by sample. Red, relatively high expression; blue, relatively low expression. **(C)** Number of differentially expressed mRNA showing up- (red) or down- (blue) regulation during development. Yellow, total number of differentially expressed mRNAs between any two stages. **(D)** Top 20 enriched GO terms for the differentially expressed mRNA between the 8- and 16-cell stage.

processes: from the oocyte to 8-cell stage and from the 16-cell to blastocyst stage. Two other minor ZGAs were found to occur between the oocyte and the 2-, 4-, and 8-cell stages, and between the morula and the blastocyst stage (**Figure 2A**, **B**). Moreover, 10,197 differentially expressed mRNAs were identified, and the largest change was also observed between 8-cell and 16-cell in two consecutive comparison groups (**Figure 2C**). Functional analysis of these differentially expressed mRNAs was enriched in 110 GO terms, including "metabolic," "binding," and "biosynthetic processes," as well as "enzymatic activity," such as "cell part," "cellular macromolecule metabolic process," "cellular biosynthetic process," "ribonucleotide binding," "phosphoprotein phosphatase activity" and other terms. This stratification indicated that goat ZGA occurs between the 8- and 16-cell stages (**Figure 2D**).

Genomic Structural Features of Goat lncRNAs

CNCI, CPC, and PFAM were used to remove potential encoded transcripts after a highly stringent filtering pipeline was applied (**Figure S1A**). A final total of 99,621 putative lncRNAs were retained (**Figure S1B**). Most of these lncRNAs (97.8%) were found to be distributed on all chromosomes except for the Y chromosome (**Table S5**). We further found that goat chromosomes 1, 2, 7 and 10 produce more lncRNAs (> 4500) than any of the others (**Figure 3A**). The identified lncRNAs were mainly divided into three categories: lincRNA, antisense lncRNA, and intronic lncRNA. Among them, intronic lncRNA was the most abundant, accounting for 65.3%, followed by lincRNA (24.9%) (**Figure 3B**). We speculated from this that these 4 chromosomes make the major contribution to the role of the lncRNAs in oocytes and pre-implantation cells growth. Combining multiple structural features to maximize our understanding of lncRNA and mRNA functions is important. The lncRNAs have an average length of 724.75 bp, which is shorter than the average protein-coding transcript length of 2872.80 bp in goat (**Figure 3C**). In addition, the lncRNAs in our current dataset were shorter than the protein-coding genes in terms of the ORF length (mean 93.78 bp vs. 520.22 bp, respectively) (**Figure 3D**).

Dynamic Expression of Differentially Expressed lncRNAs

We examined the differential expression of lncRNAs (P-adj < 0.05) between all stages of goat pre-implantation

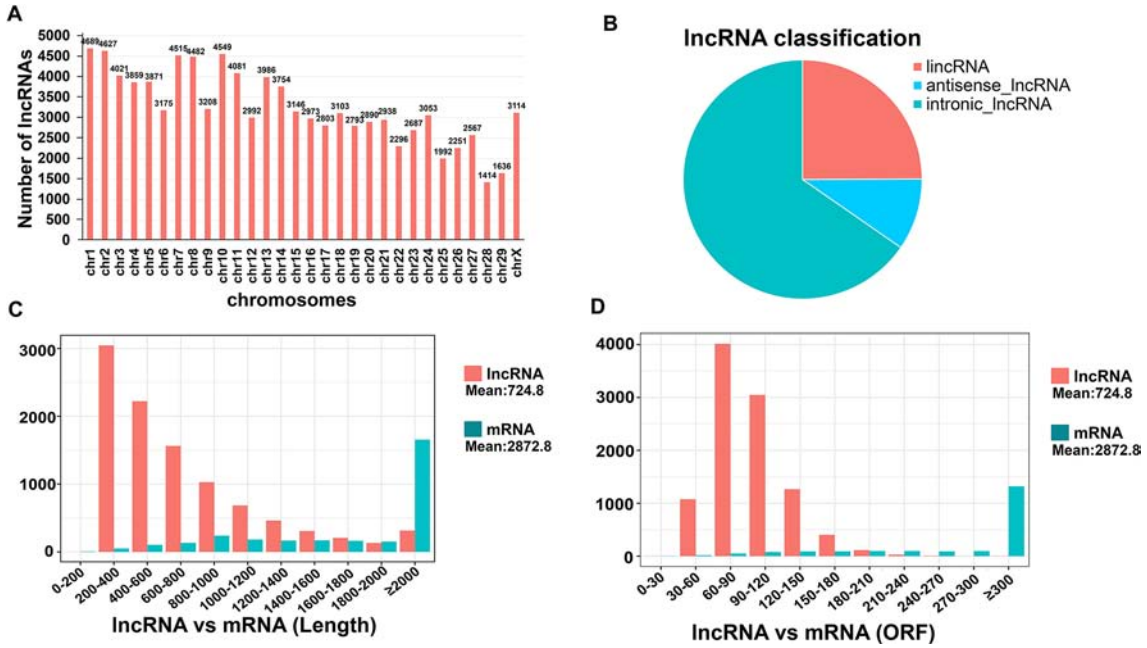


FIGURE 3 | Chromosome distribution and characteristics of lncRNAs in goat pre-implantation development. **(A)** Distribution of all identified lncRNAs in goat chromosomes. **(B)** Classification of lncRNAs. **(C–D)** Transcript length and open reading frame (ORF) length distribution of transcripts for all lncRNAs and mRNAs in goat skeletal muscle. Orange, lncRNA; blue, mRNA.

development. We identified 5160 differentially expressed lncRNAs (DELs) in these seven stages (**Figure 4A**, **Table S6**). In an unbiased hieratical clustering of these DELs, the 16-cell stage produced the largest differences from the other stages, which also confirmed the time of the major ZGA occurrence (**Figures 4A**, **B**). Interestingly, the gene expression profiles between the 2-, 4-, and 8-cell in the goat were similar. Differing from the protein-encoding transcripts, the clustering of DELs revealed that oocytes were separated from the first two cleavage events (2, 4, 8-cell stages) and the 16-cell stage was separated from the morula (**Figure 4B**). Hence, we mainly focused on the DELs between the 8- and 16-cell stages and between the oocytes and 2-cell stage.

In the major ZGA event in the goat pre-implantation stage, 905 DELs were found to be generated between the 8- and 16- cell stages, of which 780 were up-regulated and 125 were down-regulated. These DELs were enriched (P-adj < 0.05) in 24 GO terms, such as "G-protein coupled receptor activity," "G-protein coupled receptor signaling pathway," "transmembrane signaling receptor activity," and others (**Figure 4C**, **Table S7**). The minor ZGA from the oocyte to the 2-cell stage produced 148 DELs, 34 of which were up-regulated and 114 were down-regulated. Functional analysis of these two transformation stages included the "G-protein coupled receptor signaling pathway," "signaling receptor activity," "cell surface receptor signaling pathway," and others (**Figure 4D**, **Table S8**). Overall, these RNA-seq data provided an *in vivo* overview of the role of lncRNAs in ZGA waves in the goat pre-implantation stages.

WGCNA Revealing the Role of the DELs in the Developmental Transformation Leading to Pre-Implantation in the Goat

There has been no prior study describing the expression profiling of lncRNAs during goat oocyte and pre-implantation development. In addition, little functional research on these lncRNAs has been reported. To investigate the potential role of DELs in pre-implantation development, WGCNA was performed on 4761 DELs that had been filtered (FPKM > 0.01 during at least one developmental stage) and correlation analysis was conducted on the obtained modules (**Table S9**). This analysis revealed that goat DELs prior to implantation can be divided into 15 modules (denoted in the figure using different colors), 14 of which were highly correlated (correlation > 0.6, P-value < 0.05) with a specific developmental stage (**Figure 5A**, **Figure S2**). Interestingly, each preimplantation period had corresponding high expression modules. Moreover, six lncRNAs were randomly identified from stage-specific modules by qRT-PCR analysis (**Figure 5B**).

To explore DEL functions in the goat pre-implantation period, GO terminology enrichment analysis was performed for the different aforementioned modules. Interestingly, our analyses of the functions in these modules revealed a sequential progression of stage-specific core genetic networks (**Table S10**). Initially, the functional enrichment of oocyte modules (blue and salmon) included "transposase activity," "transposition," "DNA-mediated, fat cell differentiation," and others (**Figure 6A**). The functional processes migrated from "protein insertion into membrane," "DNA topoisomerase II activity," and others at the 2-cell (gray) stage, to "cell projection assembly," "cellular

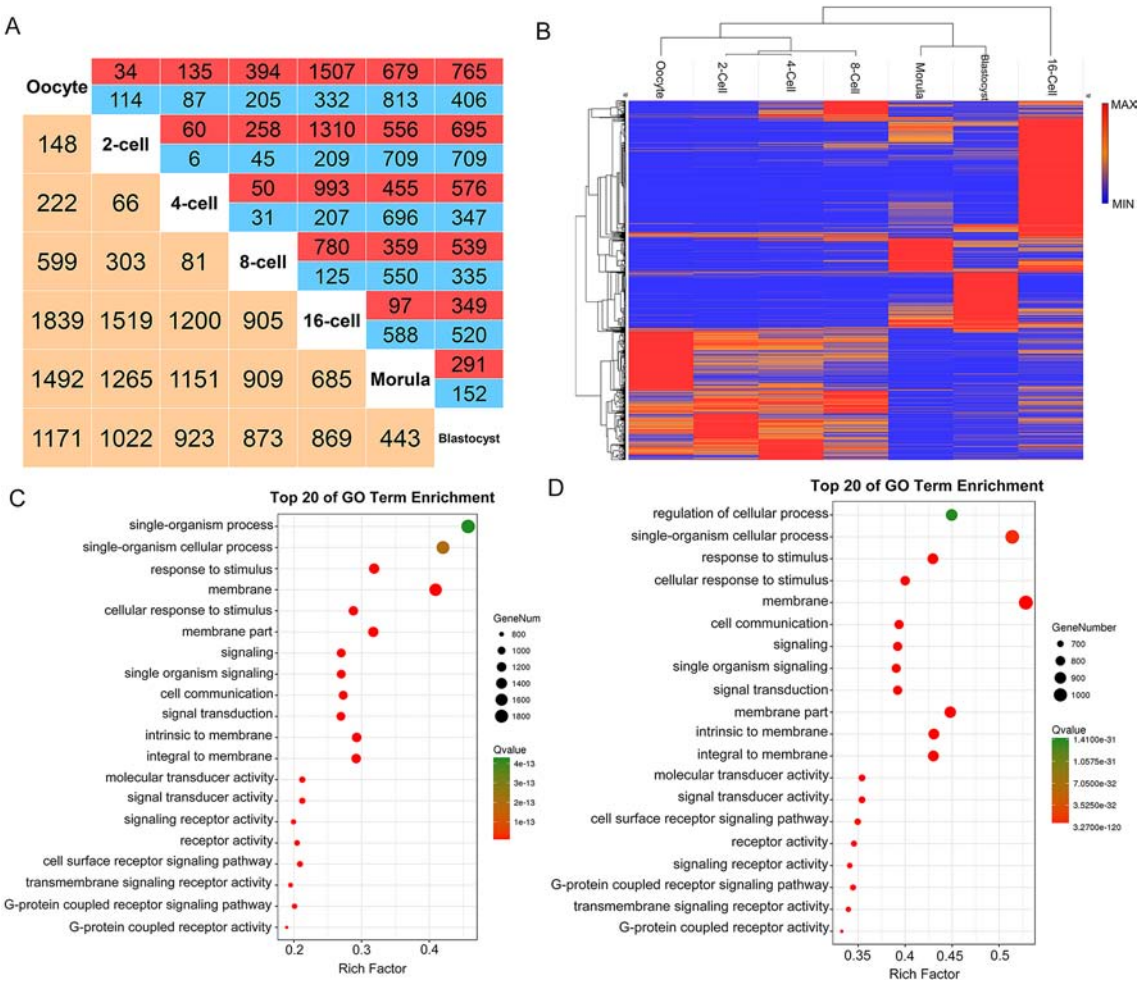


FIGURE 4 | LncRNA expression patterns during pre-implantation development in the goat. **(A)** Number of differentially expressed lncRNAs (DELs) showing up- (red) or down- (blue) regulation during development. Yellow, total number of DELs between any two stages. **(B)** Hierarchical clustering heat map of DELs by samples. Red, relatively high expression; blue, relatively low expression. **(C–D)** Top 20 enriched GO terms for the DELs between the 8- and 16-cell stage, and between the oocyte and 2-cell stage, respectively. Green, relatively high expression; red, relatively low expression.

developmental process," and others in the 4-cell (pink) stage, and then to "translation release factor activity," "translation termination factor activity," and others at the 8-cell (black) stage (**Figures 6B–F**). Functional analysis of the 16-cell stage goat modules (tan, purple, turquoise, and yellow), which occurs after the major ZGA, revealed the enrichment of 317 GO terms, including "dephosphorylation," "Ras GTPase binding," "small GTPase binding," and others (**Figure 6E**). The two other distinct major stages included "phosphoric ester hydrolase activity," "stem cell factor receptor binding," and others in the morula stage, and "protein serine/threonine kinase activity," "protein binding involved in protein folding" and others at the blastocyst stage (**Figure 6G**). Our current data thus provide the first comprehensive lncRNAs analysis of oocytes and pre-implantation stages in the goat.

To further identify lncRNAs that may play important regulatory roles in these core genetic networks, we screened the lncRNAs with the top five of "degree" as hub-lncRNAs based on the lncRNA-mRNA networks (**Figure 6**, **Figure S3**). Interestingly, most of the lncRNAs were aggregated at the 16-cell stage, which occurs after the major ZGA and produces 49919 lncRNA-mRNA pairs. Moreover, target genes for hub-lncRNAs have been identified as important participants in mammalian pre-implantation development (Pasternak et al., 2016; Daldello et al., 2019). For example, *BTG anti-proliferation factor 4* (*BTG4*) was targeted by hub-lncRNAs in goat oocyte high correlation modules, including XLOC_1684819, XLOC_2068075, and XLOC_601889 (**Table S11**). *Cyclin B2* (*CCNB2*) was also targeted by XLOC_1684819 in the oocyte stage (**Table S11**). Moreover, the top 5 hub-lncRNAs in the 6-cell stage goat all target *activating transcription factor 1*(*ATF1*), which has proved to be one of the key regulators of the ZGA (**Table S11**). These results indicate that the hub-lncRNAs we identified in our current WGCNA may have a critical regulatory role in the pre-implantation developmental stage of the goat.

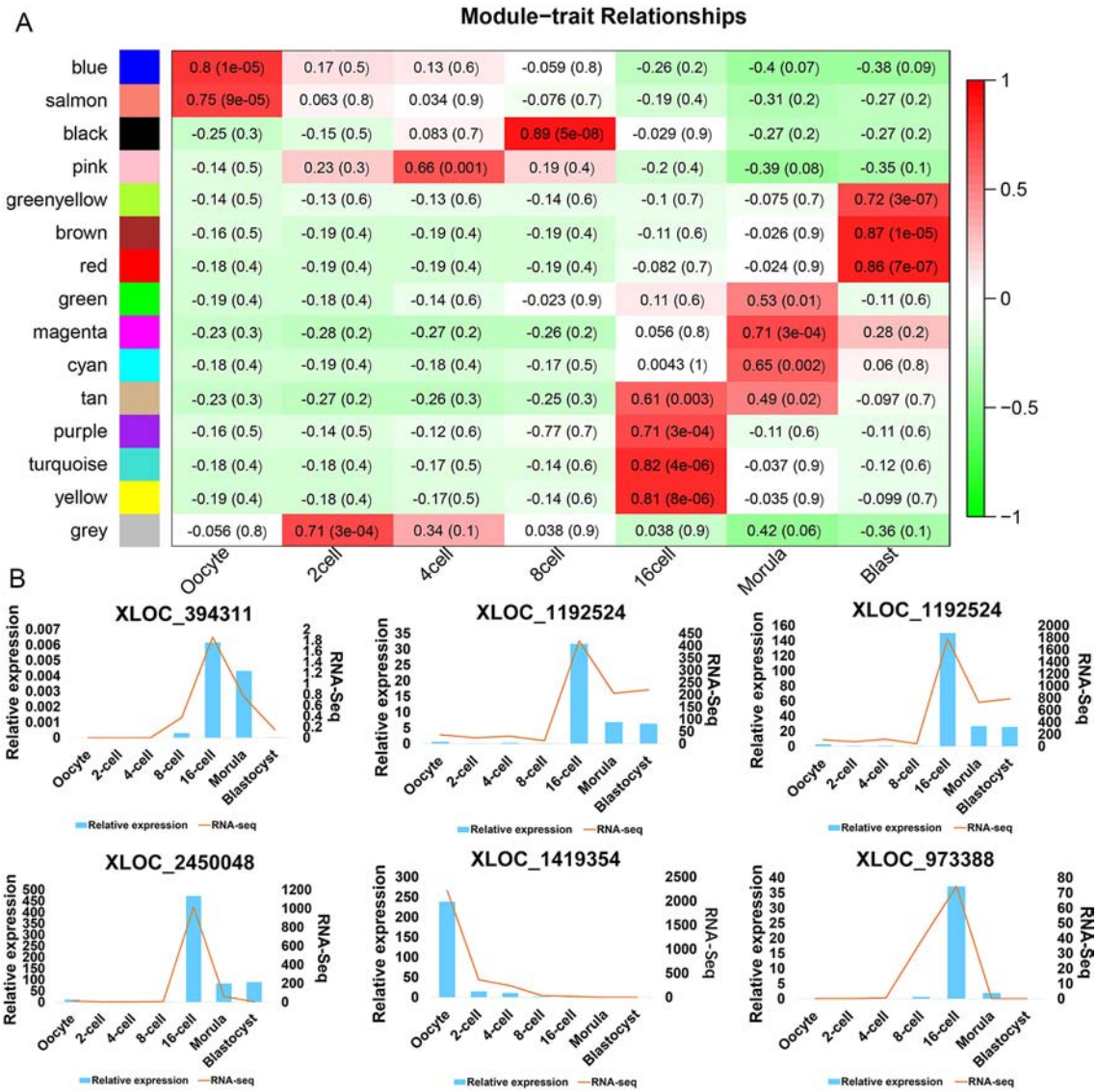
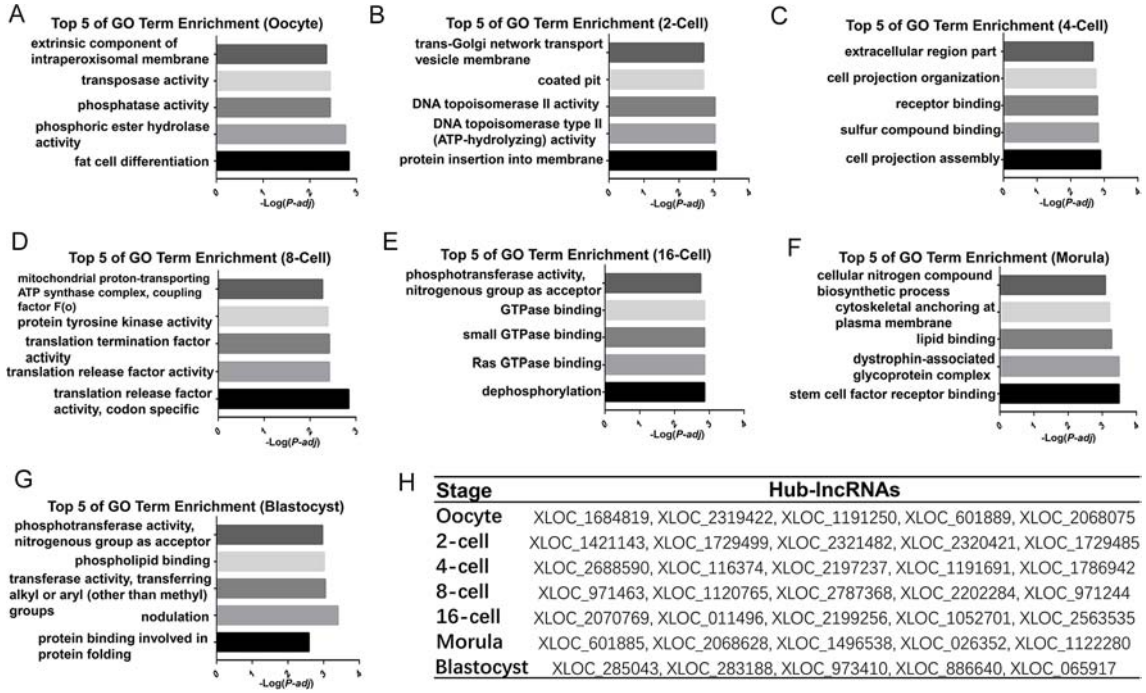


FIGURE 5 | LncRNA expression modules determined by WGCNA **(A)** Hierarchical clustering heat map of DELs (with an FPKM > 0.01 in at least one sample during the seven stages). **(B)** qPCR (bar chart, blue) and RNA-seq expression (line chart, orange) validation of the indicated lncRNAs.

DISCUSSION

The major ZGA event is the first important step in the successful initiation of mammalian pre-implantation as it results in the formation of implantable cells. This process is highly dynamic and complex, and an appropriate ZGA is essential for the normal development of the embryo (Wong et al., 2010). However, the timing of ZGA occurrences varied from species to species (Cao et al., 2014; Boroviak et al., 2018). Notably, no comprehensive lncRNA datasets have been available previously for goat pre-implantation stages. In our present study, RNA-Seq was used to analyze the transcriptome and lncRNA profiles during goat pre-implantation. The major ZGA in goat development was found in our present experiments to occur in the 8- to 16- cell stages. This was in contrast to the recent findings by Deng et al. (2018), which reported that the 8-cell stage goat stopped developing in *in vitro* developmental block cultures and showed that ZGA occurred in the 4- and 8-cell stages *in vitro*. However, other studies had shown that the timing of ZGA onset in pre-implantation cells was different between *in vitro* and *in vivo* (Misirlioglu et al., 2006; Graf et al., 2014).

We additionally explored the role of the DELs (n = 5,160) in the pre-implantation process in the goat. The functions of these molecules were identified in major and minor ZGA events that occur in the 8- to 16-cell stage and from the oocyte to 2-cell stage, respectively. The lncRNAs involved in both ZGAs were found to be enriched in "G-protein coupled receptor activity," "G-protein coupled receptor signaling pathway," and other functions related to membrane transduction and biological regulation. It is well known that G-protein coupled receptors play a key role in cell



H

Stage	Hub-lncRNAs
Oocyte	XLOC_1684819, XLOC_2319422, XLOC_1191250, XLOC_601889, XLOC_2068075
2-cell	XLOC_1421143, XLOC_1729499, XLOC_2321482, XLOC_2320421, XLOC_1729485
4-cell	XLOC_2688590, XLOC_116374, XLOC_2197237, XLOC_1191691, XLOC_1786942
8-cell	XLOC_971463, XLOC_1120765, XLOC_2787368, XLOC_2202284, XLOC_971244
16-cell	XLOC_2070769, XLOC_011496, XLOC_2199256, XLOC_1052701, XLOC_2563535
Morula	XLOC_601885, XLOC_2068628, XLOC_1496538, XLOC_026352, XLOC_1122280
Blastocyst	XLOC_285043, XLOC_283188, XLOC_973410, XLOC_886640, XLOC_065917

FIGURE 6 | Functional prediction of the high correlation modules involved in pre-implantation development in the goat **(A–G)** Bar plots showing the top 5 GO enrichment terms in the high correlation modules of each developmental stage in the goat. The length of the bars indicates the significance (-$\log_{10}$ transferred *P*-value). **(H)** Identified top5 degree hub-lncRNAs corresponding to the development stage.

self-renewal, differentiation, and signal transduction (Kobayashi et al., 2010; Rutz and Klein, 2015). Our current findings thus revealed that the lncRNAs regulate the cell membrane and its receptors during the ZGA to transduce extracellular physical and chemical signals, and thus play a role in the physiological activities of this process.

We further explored high-correlation lncRNAs at each goat stage and identified each stage of hub-lncRNA according to the lncRNA-mRNA network. The functions of the lncRNAs in these modules migrated from "transposase activity" in oocytes, to "protein insertion into membrane" during the 2-cell stage, to "cell projection assembly" at the 4- cell stage and "translation release factor activity" at the 8-cell stage, to"dephosphorylation" at the 16-cell stage, to "phosphoric ester hydrolase activity" in the morulae, and finally to "protein serine/threonine kinase activity" in the blastocyst. The transformation of the target gene enrichment function at each stage reveals the previously little-known developmental planning role of lncRNAs in goat pre-implantation cells. Furthermore, based on the lncRNA-mRNA networks in the modules and their high correlation with specific development stages, we screened for hub-lncRNAs that are potential key regulators of each pre-implantation stage during goat pre-implantation development. For example, *BTG4,* targeted by XLOC_1684819, XLOC_2068075, and XLOC_601889 lncRNAs, is a meiotic cell cycle-coupled maternal-zygotic-transition licensing factor in oocytes (Pasternak et al., 2016). *BTG4-null* female mice produce morphologically normal oocytes but are infertile due to early developmental arrest (Yu et al., 2016). *CCNB2,* targeted by XLOC_1684819, was also required for progression through meiosis in the oocyte stage (Daldello et al., 2019). Additionally, top 5 hub-lncRNAs in the 16-cell stage goat all target *ATF1*, which might prove to be one of the key regulators of the major ZGA. The presence of activated *ATF1* within the mouse nucleus at the time of ZGA indicates that this transcription factor is a priority target and a key regulator of this event (Jin and O'Neill, 2014; Orozco-Lucero et al., 2017). The DELs that highly correlate with each stage of pre-implantation transformation provides a guide for future studies of the lncRNAs that function in goat pre-implantation development. In addition, the identification of hub-lncRNAs in *in vivo* pre-implantation cells provides a valuable resource for further study of the molecular mechanisms underlying pre-implantation development.

CONCLUSION

The *in vivo* transcriptome of metaphase II oocytes, 2-, 4-, 8-, and 16-cell stage cells, and the morula and blastocyst in the goat were analyzed by RNA-Seq. The expression profile of the protein-coding genes indicates that the main ZGA occurs between the 8- and 16-cell stages. The expression profile of the DELs was also verified and these molecules play an important role in the transport and transduction of various substances during the ZGA. In addition, we described the functional continuity of the core genetic network specific for goat pre-implantation developmental stages and identify five hub-lncRNAs in each

stage. The role of lncRNA in goat oocytes and pre-implantation development had not been fully elucidated, and our current findings provided valuable resources for future research.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in the SRA database. The Accession number is PRJNA543590.

ETHICS STATEMENT

This study was carried out in accordance with the principles of the Basel Declaration and recommendations of the Guide for the Care and Use of Laboratory Animals (http://grants1.nih.gov/grants/olaw/references/phspol.htm). The protocol was approved by the ethics committee of Anhui Agricultural University under permit No. AHAU20101025.

AUTHOR CONTRIBUTIONS

Y-hL and Y-sL conceived the study, and developed hypothesis and research question. QZ, M-hS, and HW participated in the collection and processing of materials. QZ and Y-hL analyzed and interpreted the patient data. QZ carried out qRT-PCR and analyzed data. QZ, Y-hL, and Y-sL participated in the drafting and revision of the manuscript. Y-hZ, HW, Y-hL, M-xC, Y-hM, F-gF, and L-nX contributed to the writing of the manuscript. All authors reviewed the manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (31772566), the State Scholarship Fund of China Scholarship Council (201808340031), the Key Research Projects of Natural Science in Anhui Colleges and Universities (KJ2017A334), and the Agricultural Science and Technology Innovation Program of China (ASTIP-IAS13).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01040/full#supplementary-material

SUPPLEMENTAL FIGURE 1 | Goat lncRNAs identification pipeline. **(A)** Overview of goat lncRNAs identification pipeline. **(B)** Venn diagram presentation for prediction of coding potential using three software identifications, including CPC analysis, CNCI analysis, and pfam protein domain analysis.

SUPPLEMENTAL FIGURE 2 | Hierarchical cluster tree of all DELs modules. **(A)** Hierarchical cluster tree of all DELs modules in the goat embryo. Modules correspond to the branch and are denoted by the color strips under the tree.

SUPPLEMENTAL FIGURE 3 | Major lncRNA-mRNA subnetworks of high correlation modules in each preimplantation stage. **(A–G)** The red circle represents lncRNAs and the orange V shape represents co-expression mRNAs, and the size is expressed in degrees.

SUPPLEMENTAL TABLE S1 | Summary information of transcriptome.

SUPPLEMENTAL TABLE S2 | Result files for candidate lncRNAs by Cufflinks.

SUPPLEMENTAL TABLE S3 | Primer pairs used to qRT-PCR amplification.

SUPPLEMENTAL TABLE S4 | All identified mRNAs during the seven stages.

SUPPLEMENTAL TABLE S5 | All identified lncRNAs during the seven stages.

SUPPLEMENTAL TABLE S6 | Different expressed lncRNAs (DELs) of across the seven stages.

SUPPLEMENTAL TABLE S7 | The GO enriched term of DELs (8- vs 16-cell stage).

SUPPLEMENTAL TABLE S8 | The GO enriched term of DELs (oocytes vs 2-cell embryos).

SUPPLEMENTAL TABLE S9 | Number of DELs (FPKM > 0.01 at least one sample during seven stages) contained in the stage high correlation modules.

SUPPLEMENTAL TABLE S10 | GO enriched terms for the corresponding stages of the high correlation modules.

SUPPLEMENTAL TABLE S11 | Target genes for each module of hub-lncRNAs.

REFERENCES

Agliano, F., Rathinam, V. A., Medvedev, A. E., Vanaja, S. K., and Vella, A. T. (2019). Long Noncoding RNAs in Host-Pathogen Interactions. *Trends immunol.* 40(6):492–510. doi: 10.1016/j.it.2019.04.001

Baguisi, A., Behboodi, E., Melican, D. T., Pollock, J. S., Destrempes, M. M., Cammuso, C, et al. (1999). Production of goats by somatic cell nuclear transfer. *Nat. biotechnol.* 17(5):456–461. doi: 10.1038/8632

Bao, Z., Lin, J., Ye, L., Zhang, Q., Chen, J., Yang, Q., et al. (2016). Modulation of Mammary Gland Development and Milk Production by Growth Hormone Expression in GH Transgenic Goats. *Front. Physiol.* 7:278. doi: 10.3389/fphys.2016.00278

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones., S., et al. (2004). The Pfam protein families database. *Nucleic Acids Res.* 32(Database issue):D138–D141. doi: 10.1093/nar/gkh121

Bissiere, S., Gasnier, M., Alvarez, Y. D., and Plachta, N. (2018). Cell Fate Decisions During Preimplantation Mammalian Development. *Curr. Top. Dev. Biol.* 128:37–58. doi: 10.1016/bs.ctdb.2017.11.001

Boroviak, T., Stirparo, G. G., Dietmann, S., Hernando-Herraez, I., Mohammed, H., Reik, W., et al. (2018). Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development (Cambridge, England)*. 145(21). doi: 10.1242/dev.167833

Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., et al. (2014). Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC genomics*. 15:4. doi: 10.1186/1471-2164-15-4

Cuffcompare program of Cufflinks [cited 2018 November 15]. Available from: http://cole-trapnell-lab.github.io/cufflinks/cuffcompare/index.html#transfrag-class-codes

Cytoscpace software [cited 2018 December 1st]. Available from: https://cytoscape.org/download.html

Daldello, E. M., Luong, X. G., Yang, C. R., Kuhn, J., and Conti, M. (2019). Cyclin B2 is required for progression through meiosis in mouse oocytes. *Development (Cambridge, England)* 146(8). doi: 10.1242/dev.172734

Deng, M., Liu, Z., Ren, C., Zhang, G., Pang, J., Zhang, Y., et al. (2018). Long noncoding RNAs exchange during zygotic genome activation in goat. *Biol. Reprod.* 99(4):707–717. doi: 10.1093/biolre/ioy118

Deng, R., Han, C., Zhao, L., Zhang, Q., Yan, B., Cheng, R., et al. (2018). Identification and characterization of ERV transcripts in goat embryos. *Reproduction (Cambridge, England)*. doi: 10.1530/REP-18-0336

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279–D285.doi: 10.1093/nar/gkv1344

Graf, A., Krebs, S., Zakhartchenko, V., Schwalb, B., Blum, H., and Wolf, E. (2014). Fine mapping of genome activation in bovine embryos by RNA sequencing. *PNAS* 111(11):4139–4144. doi: 10.1073/pnas.1321569111

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, Y., Na, R., et al. (2016). Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). *Sci. Rep.* 6:36372. doi: 10.1038/srep36372

Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28(5):503–510. doi: 10.1038/nbt.1633

Hamazaki, N., Uesaka, M., Nakashima, K., Agata, K., and Imamura, T. (2015). Gene activation-associated long noncoding RNAs function in mouse preimplantation development. *Development (Cambridge, England)*. 142(5):910–920. doi: 10.1242/dev.116996

Jin, XL., and O'Neill, C. (2014). The regulation of the expression and activation of the essential ATF1 transcription factor in the mouse preimplantation embryo. *Reproduction (Cambridge, England)*. 148(2):147–157. doi: 10.1530/REP-13-0535

Karlic, R., Ganesh, S., Franke, V., Svobodova, E., Urbanova, J., Suzuki, Y., et al. (2017). Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 24(2):129–141. doi: 10.1093/dnares/dsw058

Kobayashi, N. R., Hawes, S. M., Crook, J. M., and Pebay, A. (2010). G-protein coupled receptors in stem cell self-renewal and differentiation. *Stem Cell Rev.* 6(3):351–366. doi: 10.1007/s12015-010-9167-9

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35(Web Server issue):W345–W349. doi: 10.1093/nar/gkm391

Kurian, L., Aguirre, A., Sancho-Martinez, I., Benner, C., Hishida, T., Nguyen, TB., et al. (2015). Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. *Circulation*. 131(14):1278–1290. doi: 10.1161/CIRCULATIONAHA.114.013303

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics (Oxford, England)*. 24(5):719–720. doi: 10.1093/bioinformatics/btm563

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4):357–359. doi: 10.1038/nmeth.1923

Lee, M. T., Bonneau, A. R., and Giraldez, A. J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* 30:581–613. doi: 10.1146/annurev-cellbio-100913-013027

Li, L., Chen, Y., Nie, L., Ding, X., Zhang, X., Zhao, W., et al. (2019). MyoD-induced circular RNA CDR1as promotes myogenic differentiation of skeletal muscle satellite cells. *Biochim. Biophys. Acta Gene. Regul. Mech.* 1862(8):807–821. doi: 10.1016/j.bbagrm.2019.07.001

Li, Y. P., Duan, F. F., Zhao, Y. T., Gu, K. L., Liao, L. Q., Su, H. B., et al. (2019). A TRIM71 binding long noncoding RNA Trincr1 represses FGF/ERK signaling in embryonic stem cells. *Nature communications*. 10(1):1368. doi: 10.1038/s41467-019-08911-w

Lokken, A. A., and Ralston, A. (2016). The Genetic Regulation of Cell Fate During Preimplantation Mouse Development. *Current topics in developmental biology* 120:173–202. doi: 10.1016/bs.ctdb.2016.04.006

Ma, L. B., He, X. Y., Wang, F. M., Cao, J. W., and Cheng, T. (2014). The development and expression of pluripotency genes in embryos derived from nuclear transfer and in vitro fertilization. *Zygote (Cambridge, England)*. 22(4):540–548. doi: 10.1017/S0967199413000129

Misirlioglu, M., Page, G. P., Sagirkaya, H., Kaya, A., Parrish, J. J., First, N. L., et al. (2006). Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. *Proc. Natl. Acad. Sci. U.S.A.* 103(50):18905–18910. doi: 10.1073/pnas.0608247103

Morpheus [cited 2019 April 12]. Available from: https://software.broadinstitute.org/morpheus

Omicshare tools [cited 2019 April 5th]. Available from: www.omicshare.com/tools

Orozco-Lucero, E., Dufort, I., and Sirard, M. A. (2017). Regulation of ATF1 and ATF2 transcripts by sequences in their 3' untranslated region in cleavage-stage cattle embryos. *Mol. Reprod. Dev.* 84(4):296–309. doi: 10.1002/mrd.22785

Pasternak, M, Pfender, S, Santhanam, B, and Schuh, M. (2016). The BTG4 and CAF1 complex prevents the spontaneous activation of eggs by deadenylating maternal mRNAs. *Open Biol.* 6(9). doi: 10.1098/rsob.160184

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40(Database issue):D290–D301. doi: 10.1093/nar/gkr1065

Rutz, C., Klein, W, and Schulein R. (2015). N-Terminal Signal Peptides of G Protein-Coupled Receptors: Significance for Receptor Biosynthesis, Trafficking, and Signal Transduction. *Prog. Mol Biol Transl. Sci.* 132:267–287. doi: 10.1016/bs.pmbts.2015.03.003

SPSS Statistics [cited 2018 May 6th]. Available from: https://www.ibm.com/products/spss-statistics

Sun, M., and Kraus, W. L. (2013). Minireview: Long noncoding RNAs: new "links" between gene expression and cellular outcomes in endocrinology. *Mol. Endocrinol (Baltimore, Md)*. 27(9):1390–1402. doi: 10.1210/me.2013-1113

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 41(17):e166. doi: 10.1093/nar/gkt646

The Capra hircus reference gene model annotation file [cited 2018 October 27]. Available from: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1/GCF_001704415.1_ARS1_genomic.gff.gz

The Capra hircus reference genome model annotation file [cited 2018 October 27]. Available from: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1/GCF_001704415.1_ARS1_genomic.fna.gz

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*. 25(9):1105–1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5):511–515. doi: 10.1038/nbt.1621

Vallot, C., Patrat, C., Collier, A. J., Huret, C., Casanova, M., Liyakat Ali, T. M., et al. (2017). XACT Noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell stem cell*. 20(1):102–111. doi: 10.1016/j.stem.2016.10.014

Wong, C. C., Loewke, K. E., Bossert, N. L., Behr, B., De Jonge, C. J., Baer, T. M., et al. (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat. Biotechnol.* 28(10):1115–1121. doi: 10.1038/nbt.1686

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*. 500(7464):593–597doi: 10.1038/nature12364

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20(9):1131–1139. doi: 10.1038/nsmb.2660

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11(2):R14. doi: 10.1186/gb-2010-11-2-r14

Yu, C., Ji, S. Y., Sha, Q. Q., Dang, Y., Zhou, J. J., Zhang, Y. L., et al. (2016). BTG4 is a meiotic cell cycle-coupled maternal-zygotic-transition licensing factor in oocytes. *Nat. Struct.Mol. Biol.* 23(5):387–394. doi: 10.1038/nsmb.3204

Yuan, S., Schuster, A., Tang, C., Yu, T., Ortogero, N., Bao, J., et al. (2016). Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development (Cambridge, England).* 143(4):635–647. doi: 10.1242/dev.131755

Zhong, L., Mu, H., Wen, B., Zhang, W., Wei, Q., Gao, G., et al. (2018). Long non-coding RNAs involved in the regulatory network during porcine pre-implantation embryonic development and iPSC induction. *Scientific reports* 8(1):6649. doi: 10.1038/s41598-018-24863-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer ZH declared a past co-authorship with one of the authors YM to the handling editor.

Copyright © 2019 Ling, Zheng, Li, Sui, Wu, Zhang, Chu, Ma, Fang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 October 2019
doi: 10.3389/fgene.2019.01070

Gene Expression and Fatty Acid Profiling in *Longissimus thoracis* Muscle, Subcutaneous Fat, and Liver of Light Lambs in Response to Concentrate or Alfalfa Grazing

*Elda Dervishi[1], Laura González-Calvo[2], Mireia Blanco[2], Margalida Joy[2], Pilar Sarto[2], R. Martin-Hernandez[3], Jose M. Ordovás[4], Magdalena Serrano[5] and Jorge H. Calvo[2,6]**

[1] Livestock Gentec, University of Alberta, Edmonton, AB, Canada, [2] Unidad de Producción y Sanidad Animal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA)-Instituto Agroalimentario de Aragón (IA2) (CITA-Universidad de Zaragoza), Zaragoza, Spain, [3] Precision Nutrition and Obesity, IMDEA-Alimentación, Madrid, Spain, [4] Jean Mayer-USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States, [5] Departamento de Mejora Genética Animal, INIA, Madrid, Spain, [6] ARAID, Zaragoza, Spain

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Andrea Serra, University of Pisa, Italy
Paula Alexandra Lopes, University of Lisbon, Portugal

***Correspondence:**
Jorge H. Calvo
jhcalvo@aragon.es

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *09 April 2019*
Accepted: *04 October 2019*
Published: *31 October 2019*

Citation:
Dervishi E, González-Calvo L, Blanco M, Joy M, Sarto P, Martin-Hernandez R, Ordovás JM, Serrano M and Calvo JH (2019) Gene Expression and Fatty Acid Profiling in Longissimus thoracis Muscle, Subcutaneous Fat, and Liver of Light Lambs in Response to Concentrate or Alfalfa Grazing. Front. Genet. 10:1070. doi: 10.3389/fgene.2019.01070

A better understanding of gene expression and metabolic pathways in response to a feeding system is critical for identifying key physiological processes and genes associated with polyunsaturated fatty acid (PUFA) content in lamb meat. The main objective of this study was to investigate transcriptional changes in *L. thoracis* (LT) muscle, liver, and subcutaneous fat (SF) of lambs that grazed alfalfa (ALF) and concentrate-fed (CON) slaughtered at 23 kg and using the Affymetrix Ovine Gene 1.1 ST whole-genome array. The study also evaluated the relationship between meat traits in LT muscle, including color, pigments and lipid oxidation during 7 days of display, α-tocopherol content, intramuscular fat (IMF) content and the fatty acid (FA) profile. Lambs that grazed on alfalfa had a greater α-tocopherol concentration in plasma than CON lambs ($P < 0.05$). The treatment did not affect the IMF content, meat color or pigments ($P > 0.05$). Grazing increased the α-tocopherol content ($P < 0.001$) and decreased lipid oxidation on day 7 of display ($P < 0.05$) in LT muscle. The ALF group contained a greater amount of conjugated linoleic acid (CLA), C18:3 n–3, C20:5 n–3, C22:5 n–3, and C22:6 n–3 than did the CON group ($P < 0.05$). We identified 41, 96 and four genes differentially expressed in LT muscle, liver, and subcutaneous fat, respectively. The most enriched biological processes in LT muscle were skeletal muscle tissue development, being the genes related to catabolic and lipid processes downregulated, except for *CPT1B*, which was upregulated in the ALF lambs. Animals grazing alfalfa had lower expression of desaturase enzymes in the liver (*FADS1* and *FADS2*), which regulate unsaturation of fatty acids and are directly involved in the metabolism of n–3 PUFA series. The results found in the current study showed that ingesting diets richer in n–3 PUFA might have negative effects on the *de novo* synthesis of n–3 PUFA by downregulating the *FADS1* and *FADS2* expression. However, feeding diets poorer in n–3 PUFA can promote fatty acid desaturation, which makes these two genes attractive candidates for altering the content of PUFAs in meat.

Keywords: concentrate, alfalfa, microarray, ovine, muscle, subcutaneous fat, liver, meat quality

INTRODUCTION

Public health policies recommend an increase in the intake of the n−3 polyunsaturated fatty acid (PUFA) series due to the positive impact these molecules have on human health. In addition, a decrease in the consumption of trans-fatty acids and saturated fatty acids (SFAs) is recommended because they have been associated with increased cholesterol levels (Takeuchi and Sugano, 2017; Zhu et al., 2019). Other fatty acids, such as conjugated linoleic acids (CLAs), have also received increasing attention because of their possible beneficial effects on human health (Lehnen et al., 2015; Lee et al., 2018).

Currently, increasing feed efficiency and producing lean meat without reducing the nutritional value of the meat are major challenges of the meat industry. The nutritional value of meat can be influenced by dietary and genetic effects (Scollan et al., 2014). Grass feeding improves eicosapentaenoic acid (20:5n−3, EPA), docosapentaenoic acid (DPA, 22:5n−3), and docosahexaenoic acid (22:6n−3, DHA) contents in muscle (Fisher et al., 2000; Dervishi et al., 2010; Dervishi et al., 2011) as forage increases the content of alpha-linolenic acid (18:3n−3), the precursor for DHA and EPA production (Kitessa et al., 2010). Diet has been shown to have a major impact on the intramuscular FA profile of the muscle of light lambs (Dervishi et al., 2010; González-Calvo et al., 2015a); grazing increases the PUFA content of the n−3 series and conjugated fatty acids (CLAs) when compared to that with concentrate feeding. In *semitendinosus* muscle, genes related to adipogenesis are upregulated in concentrate-fed lambs, whereas *CPT1B* gene expression, related to the β-oxidation process, is upregulated in grazing lambs (Dervishi et al., 2011). However, the expression of genes implicated in lipid metabolism is not similar in the *longissimus* muscle of grazing and concentrate-fed lambs (González-Calvo et al., 2015a). These results demonstrate that the diet/feeding system has a differential effect on gene expression in different animal tissues. It has also been demonstrated that fiber type composition in skeletal muscle (the relative amounts of fast versus slow twitch fibers) affected the gene expression profiles among different muscle under the same environment (Terry et al., 2018). Therefore, a better understanding of the genes and metabolic pathways in response to the feeding system is critical for identifying key physiological processes and genes associated with lipid metabolism, especially for the n−3 PUFA series. A deeper understanding of the gene regulation of n−3 levels in lamb meat may help in designing new strategies for the production of healthier meat and satisfying consumers' demand.

The combination of technologies such as fatty acids and gene profiling provides a powerful tool for discovering gene expression changes associated with meat quality traits and for discovering genes contributing to fatty acid content variation in meat. The main objective of this study was to investigate the fatty acid profile and transcriptional changes in the LT muscle, liver, and subcutaneous fat (SF) of lambs grazing on alfalfa pasture and receiving concentrate using the Affymetrix Ovine Gene 1.1 ST whole-genome array. Furthermore, we aimed to identify novel genes that may play important roles in the metabolism of PUFAs that may be associated with meat quality traits.

MATERIAL AND METHODS

Ethics Statement

All experimental procedures, including the care of animals and euthanasia, were performed in accordance with the guidelines of the European Union and Spanish regulations for the use and care of animals in research and were approved by the Animal Welfare Committee of the Centro de Investigación y Tecnología Agroalimentaria (CITA) (protocol number 2009-01_MJT). In all cases, euthanasia was performed by penetrating captive bolt followed by immediate exsanguination.

Animals and Sample Collection

Fourteen pairs of ewe-single reared male lambs of the Rasa Aragonesa breed grazed continuously during lactation in alfalfa pastures. The lambs had *ad libitum* access to a concentrate during lactation. Seven pairs of ewe-lambs were not weaned but remained grazing alfalfa with their mothers from birth until the slaughter of the lambs (23 ± 0.4 kg) (ALF group). The other seven lambs were weaned (48 ± 0.9 days of age) and then fed a basal concentrate for 24 (± 2.6) days until slaughter at 23 kg (CON group). These lambs were the same as those described in González-Calvo et al. (2017), and were reared alongside the ALF group. Lambs belonging to ALF treatment received dams' milk, fresh alfalfa (grazing) and commercial concentrate, the same that was offered to CON treatment during the experimental period. The average concentrate intake of the CON and ALF groups during the experimental period was 24.3 and 7.4 kg per lamb, respectively. The weaning weight of CON treatment animals was 11.6 ± 1.91 kg BW and the weight of the alfalfa lambs at the same moment of the weaning of CON treatment was 12.8 ± 1.35 kg BW. The ingredients, chemical composition and FA composition of the feedstuffs are shown in **Table 1**. The experimental procedures, composition of diets, management of the animals and sample details for each group are described in detail in Ripoll et al. (2013). Blood samples were obtained weekly in test tubes containing heparin from the jugular vein. Samples were centrifuged at 3,500 rpm for 20 min, and plasma was stored at −80°C until α-tocopherol and triacylglycerols (TG), cholesterol, low density lipoprotein-cholesterol (LDL-cholesterol) and high density lipoprotein-cholesterol (HDL-cholesterol) analyses.

All the lambs were slaughtered when they reached 22–24 kg of slaughter weight (SW) according to the specifications of Ternasco de Aragón Protected Geographical Indication (Regulation (EC) No. 1107/96) that stipulates that lambs must be younger than 90 days old with a SW between 22 and 24 kg. The lambs were slaughtered using EU laws in the same commercial abattoir, and the carcasses were hung by the Achilles tendon and chilled for 24 h at 4°C in total darkness. The slaughter age, slaughter weight, and growth rate of the two management strategies are presented in **Supplementary Table 1**.

Just after slaughter, a sample of the LT muscle from the 12th thoracic *vertebra*, a sample of SF between the atlas and axis *cervical vertebrae* and a sample of the liver were excised, frozen in liquid nitrogen and stored at −80°C until RNA isolation.

TABLE 1 | Ingredients and chemical composition of the feedstuffs used in the experiment.

Item	Treatment[1]	
	ALF	CON
Ingredients (%)		
Barley		40.87
Corn		14.95
Wheat		20.08
Soyabean meal		19.78
Salt		0.39
Carbonate		1.64
Mineral–vitamin mixture		1.20
Fat		1
Chemical composition[2]		
g DM/kg as fed	24.6	88.6
CP, g/kg DM	154	175
CF, g/kg DM		38
NDF, g/kg DM	326	180
ADF, g/kg DM	204	45
α-tocopherol, mg/kg DM DM	154	27[3]
FA composition (g/100 g FA)[4]		
C16:0	17.7	24.35
C18:0	3.02	2.91
C18:1 c9	4.52	25.63
C18:2 n–6	22.28	38.86
C18:3 n–3	37.93	3.14
C20:0	4.46	0.33
C22:0	3.03	0

[1]CON: commercial concentrate; ALF: unweaned lambs grazing alfalfa plus commercial concentrate.
[2]DM, dry matter; CP, crude protein; CF, crude fat; NDF, neutral detergent fibre; ADF, acid detergent fibre.
[3]As mg dl-α-tocopheryl acetate/kg DM.
[4]Fatty acid composition expressed as the percentage of total fatty acid methyl esters.

Chemical Analyses

Intramuscular Fat (IMF)

The intramuscular fat content was quantified using the Ankom procedure (AOAC, 2000) with an Ankom extractor (model XT10, Ankom Technology, New York, USA).

Fatty Acid Determination

Both muscle and feed fatty acids were determined as described in González-Calvo et al. (2015a). Feed samples were Soxhlet extracted (Sukhija and Palmquist, 1988), and muscle samples were determined according to Bligh and Dyer (1959) with the modifications described in González-Calvo et al. (2015a). The individual FA contents were expressed as weight percentages (g/100 g of FAME). The total amount of SFA, monounsaturated FA (MUFA), PUFA, n−6 PUFA and n−3 PUFA contents and their associated ratios (PUFA : SFA and n−6:n−3) were determined.

Analysis of α-Tocopherol, TG, LDL-Cholesterol, HDL-Cholesterol, and Cholesterol in Plasma

Alpha-tocopherol in plasma was determined by liquid extraction in duplicate as described in González-Calvo et al. (2015b). Triacylglycerols, cholesterol, LDL-cholesterol and HDL-cholesterol were determined using an automatic analyzer (Gernonstar, RAL, Barcelona, España). The reagent manufacturer was RAL (Técnica para el Laboratorio, S.A. Sant Joan Despí, Barcelona, Spain). The mean intra-assay coefficients of variation were 0.99–1.57%, 0.76–1.22%, 0.63–0.67%, and 0.8–1.06% for TG, cholesterol, LDL-cholesterol and HDL-cholesterol, respectively. The interassay coefficients of variation were 3.15–7.77%, 4.36–6.91%, 1.29–1.45% and 2.71–4.60% for the same metabolites.

Analysis of α-Tocopherol Concentration, TBARS and Metmyoglobin Formation in Muscle

After it was chilled, a piece of the LT muscle between the 4th and the 6th lumbar *vertebrae* was vacuum-packed and kept at −20°C in darkness until the α-tocopherol analysis. The α-tocopherol concentration was determined by liquid extraction as described in González-Calvo et al. (2015b). A portion of the loin between the 7th and the 13th thoracic *vertebrae* was used to measure the color (metmyoglobin content, MMb) and lipid oxidation analysis (thiobarbituric acid-reactive substance, TBARS), and were quantified at 7 days after being maintained in darkness at 4°C. The LT muscle color and LT intramuscular fat TBARS analysis were measured as described in González-Calvo et al. (2015b). Briefly, the relative content of metmyoglobin (MMb) was estimated by the $K/S_{572/525}$ ratio (Hunt, 1980). This ratio decreases when the MMb content increases. The TBARS analysis was performed using the procedure reported by Pfalzgraf et al. (1995). The TBARS values are expressed as milligrams of malonaldehyde (MDA) kg^{-1} of muscle.

RNA Isolation and Assessment of RNA Integrity

Total RNA was extracted from approximately 500 mg of LT muscle, SF, and liver using RNeasy Tissue mini kits (QIAGEN, Madrid, Spain) following the manufacturer's protocol. Prior to microarray analysis, RNA integrity and quality were assessed by an RNA 6000 Nano LabChip on an Agilent 2100 Bioanalyzer and quantified using a nanophotometric spectrophotometer (Implen, Madrid, Spain). All RNA integrity number (RIN) values were above 8.

Microarray Hybridization and Data Processing

RNA samples (n = 14, seven samples from each treatment) were analyzed using the Ovine Gene 1.1 ST Array Strip (Affymetrix, High Wycombe, UK). Microarray hybridization and scanning were performed at the Functional Genomics Core Facility (Institute for Research in Biomedicine, IRB Barcelona, Spain) following the recommendations of the manufacturer. Scanned images (DAT files) were transformed into intensities (CEL files) by Affymetrix GeneChip Operating Software (GCOS). The overall array intensity was normalized between arrays to correct for systematic bias in the data and to remove the impact of nonbiological influences on biological data. The imported data were analyzed at the gene level, with exons summarized to genes, using the mean expression of all the exons of a gene. Normalization was carried out with the Robust Multi-Array Average (RMA) algorithm using quantile normalization, median

polish probe summarization, and log2 probe transformation. The datasets supporting the results and discussed in this publication have been deposited in the NCBI Gene Expression Omnibus repository (Barrett et al., 2012) and are accessible through GEO Series accession numbers GSE63774 (LT muscle and SF) and GSE125661 (liver). The datasets for LT muscle and SF in CON group were previously presented in González-Calvo et al. (2017).

Validation of Microarray Data by Real-Time Quantitative PCR Analysis (RT-qPCR)

One microgram of RNA from each sample was treated with DNAse (Invitrogen, Carlsbad, CA, USA), and single-stranded cDNA was synthesized using the SuperScript®III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA), following the manufacturer's recommendations. Specific exon-spanning primers for genes were generated and confirmed for specificity using BLAST (National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/BLAST/). Before performing the real-time PCRs, a conventional PCR was performed for all genes to test the primers and to verify the amplified products. The PCR products were sequenced to confirm gene identity using an ABI Prism 3700 (Applied Biosystems, Madrid, Spain) with standard protocols. Homology searches were performed with BLAST to verify the identity of the amplified fragments. The real-time PCR was carried out in a 10 µl PCR total reaction mixture containing SYBR Green Master Mix: SYBR Premix Ex Taq II (Tli RNase H Plus, Takara, Sumalsa, Zaragoza, Spain). Reactions were run in triplicate on an ABI Prism 7500 platform (Applied Biosystem, Madrid, Spain) following the manufacturer's cycling parameters. Standard curves for each gene were generated to calculate the amplification efficiency through a 4-fold serial dilution of cDNA pooled from LT muscle, liver and SF. The efficiency (E) of PCR amplification for each gene was calculated using the standard curve method ($E = 10^{(-1/slope)}$). Two "connector samples" were replicated in all plates to remove technical variation from this source of variability. The annealing temperatures, primer concentrations, and primer sequences for GOIs (Genes of interest: *CPT1B*, *MYOD1*, *MSTN*, *ABCC4*, *IGF1R* and *PLA2G16* for LT muscle; *METTL1* for SF; and *FADS1*, *FADS2*, *ACACA*, *SCD*, *SQLE*, *IER3*, *SLC19A1* and *THRSP* for liver tissue) and reference genes (*GUSB* and *YWHAZ* for LT muscle and SF; and *RPL37*, *GUSB* and *RPL19* for liver) are described in **Supplementary Table 2**. These reference genes for LT muscle and SF were chosen because they were the most stable in these tissues in previous studies (González-Calvo et al., 2014). Five candidate reference genes (*B2M*, *YWHAZ*, *RPL37*, *RPL19*, and *GUSB*) were tested for liver tissue. Determinations of the gene expression stability of liver genes included in this study were calculated using NormFinder to select the best reference genes (Andersen et al., 2004).

Statistical Analysis

Statistical Analysis of the Performance, Concentrations of TG, LDL-Cholesterol, HDL-Cholesterol, and Cholesterol in Plasma, and Meat Quality Characteristics in LT Muscle

Statistical analysis of the performance, the plasma metabolites and lipid oxidation of LT muscle (TBARS) was performed using the SAS statistical package v. 9.3 (SAS Institute, Cary NC, USA). The concentration of analytes in plasma, lipid oxidation levels and meat color and pigments were analyzed using mixed models for repeated measurements based on Kenward-Roger's adjusted degrees of freedom solution for repeated measures including the management strategy (CON and ALF), the week/time of display and its interaction as fixed effects and the lamb as the random effect. A first-order autoregressive structure with heterogeneous variances for each date was used to model heterogeneous residual error.

The weight gain, age, weight at slaughter and IMF in LT muscle were analyzed using a general lineal model (GLM) with the treatment as a fixed factor. The content of α-tocopherol and the fatty acid profile of LT muscle were analyzed with a GLM with the treatment as a fixed factor and the slaughter age (SA) as a covariate. The results were expressed as least square means (LSM) ± the standard error (SE) values, and the differences were tested at a level of significance of 0.05 with the t statistic. The Tukey *post hoc* test was used to evaluate differences between treatments.

Microarray Gene Expression Statistical Analysis

Identiȧcation of Differentially Expressed Genes by Microarray Analysis in LT Muscle and SF

Normalized data were further analyzed using Babelomics (http://babelomics.bioinfo.cipf.es/graph.html) and MetaboAnalyst software (Xia et al., 2009). Genes showing a statistically significant value of the Limma test ($P < 0.01$) were screened out as differentially expressed between treatments. Significant genes were annotated based on similarity scores in blastn comparisons of Affymetrix transcript cluster sequences against ovine sequences in GenBank. A second method, significance analysis of microarray (SAM), was used to identify and reconfirm differentially expressed genes in ALF–CON comparisons. Details of the protocol are described in González-Calvo et al. (2017).

Multivariate Analysis of Gene Expression and Hierarchical Clustering Analysis (HCA)

Multivariate and cluster analysis was performed using MetaboAnalyst according to Xia et al. (2009). Principal components analysis (PCA) was used to cluster the samples based on the selected gene expression profile for each tissue. Hierarchical clustering analysis for gene expression was performed using all genes and only the significant genes for each tissue. Details are described in González-Calvo et al. (2017).

Statistical Analysis of Gene Expression Validated by RT-qPCR

The corresponding mRNA levels were measured and analyzed by their quantification cycle (Cq). The statistical methodology to analyze differences in the expression rate was carried out following the method proposed by Steibel et al. (2009). The mixed model fitted was as follows:

$$y_{rigkm} = TG_{gi} + P_k + b_1(IMF)_m + b_2(SA)_m + A_m + e_{rigkm}$$

where y_{rigkm} is the C_q value (transformed data taking into account $E < 2$) obtained from the thermocycler software for the gth gene

(GOIs and reference genes) from the rth well (reactions were run in triplicate) in the kth plate corresponding to the mth animal and to the ith treatment (CON and ALF); TG_{gi} is the fixed interaction among the ith treatment and the gth gene (T is the effect of the ith treatment, and G is the effect of the gth gene); P_k is the fixed effect of the kth plate; IMF_m and SA_m are the effects of intramuscular fat (only used in LT muscle tissue gene expression) and the slaughter age of the mth animal, respectively, included as covariates; A_m is the random effect of the mth animal from where samples were collected ($A_m \sim (0,\sigma^2_A)$); and e_{rigkm} is the random residual. Gene-specific residual variance (heterogeneous residual) was fitted to the gene by treatment effect ($e_{rigkm} \sim N(0, \sigma^2_{egi})$).

To test differences ($diff_{GOI}$) in the expression rate of the target genes between treatments and to obtain fold change (FC) values from the estimated TG differences, the approach suggested in Steibel et al. (2009) was used. The significance of the $diff_{GOI}$ estimates was determined with the t statistic. Additionally, asymmetric 95% confidence intervals (upper and lower) were calculated for each FC value using the standard error (SE) of $diff_{GOI}$.

Functional Annotation Analyses

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7b (Huang et al., 2008) was used to determine pathways and processes of major biological significance and importance through the Functional Annotation Cluster (FAC) tool based on the Gene Ontology (GO) annotation function. DAVID FAC analysis was performed with the gene lists obtained after SAM analysis. Medium stringency EASE score parameters were selected to indicate confident enrichment scores of functional significance and importance of the given pathways and processes investigated. An enrichment score of 1.3 was employed as the threshold for cluster significance. The ClueGO plug-in (Bindea et al. 2009) and Cytoscape program (Shannon et al., 2003) were used to group genes according to the similarity of the biological processes in which they are involved. The relationships between the n−3 PUFA series in muscle and gene expression in liver were visualized using the Metscape plugin (Karnovsky et al., 2012) in Cytoscape (Shanon et al., 2003). Thirty significant genes (out of 96) and nine (out of 13) significant compounds were mapped to KEGG IDs. The file containing the list of genes and metabolites, their fold change and P-values was loaded in Metscape to generate a compound-gene network.

RESULTS

Lambs Performance and α-Tocopherol, TG, LDL-Cholesterol, HDL-Cholesterol and Cholesterol Concentrations in Plasma

No differences were found in the slaughter weight and age (SA) or average daily gain (ADG) from birth to slaughter between treatments (**Supplementary Table 1**). Type III tests of the fixed effects (treatment and day) on blood parameters are shown in **Supplementary Table 3**.

The concentrations of α-tocopherol, cholesterol, LDL-cholesterol and HDL-cholesterol in plasma were affected by the interaction between the treatment and the day ($P < 0.05$ to < 0.01). Meanwhile, the concentration of TG in plasma was affected only by the treatment ($P < 0.01$).

Grazing animals (ALF group) had a greater concentration of α-tocopherol ($P < 0.0001$) and similar cholesterol ($P = 0.056$) contents throughout the experimental period when compared to those in the CON group (**Figures 1A**, **B**). The HDL-cholesterol content was similar between treatments except on day 8 after weaning, and the ALF group had a greater content when compared to that in the CON group (0.58 ± 0.03 mmol/l vs. 0.34 ± 0.03 mmol/l; $P < 0.05$). Similarly, ALF lambs presented a greater concentration of TG on days 8 and 28 postweaning (0.58 ± 0.03 vs. 0.67 ± 0.05 and 0.25 ± 0.03 vs. 0.3073 ± 0.07 mmol/l; $P < 0.05$). The LDL-cholesterol content was greater in the ALF group only on day 0 (equivalent to the day of weaning in the CON group) when compared to that in the CON group (0.48 ± 0.05 vs. 0.18 ± 0.05 mmol/l; $P < 0.05$) (**Figure 1**).

Meat Characteristics in *Longissimus thoracis* Muscle

Intramuscular fat content, meat color and pigments were not different between treatments (**Supplementary Table 4**; $P > 0.05$); however, ALF lambs had a greater content of α-tocopherol in LT muscle when compared with that in the CON group (2.38 ± 0.17 vs. 0.48 ± 0.17; $P < 0.05$) (**Supplementary Table 3**). Lipid oxidation was affected by the interaction between the treatment and the day of display ($P < 0.01$). Lipid oxidation was similar in the first days of display ($P > 0.05$), but ALF lambs had lower oxidation than did CON lambs on day 7 of display ($P < 0.05$).

Regarding the fatty acid profile, LT muscle from the ALF group had a greater content of capric (C10:0) and arachidic (C20:0) acids (**Table 2**; $P < 0.05$) and tended to have a greater content of margaric acid (C17:0) and stearic acid (C18:0) ($0.1 < P > 0.05$). However, the total SFA content was not different between treatments ($P > 0.05$; **Table 2**).

The treatment did not affect palmitoleic acid (C16:1; $P > 0.05$), vaccenic acid (C18:1 n−7; $P > 0.05$) and eicosenoic acid (C20:1 n−9; $P > 0.05$) content but did affect the content of oleic acid (C18:1 n−9; $P < 0.05$) and total MUFAs, which was greater in CON than in ALF lambs ($P < 0.05$; **Table 2**). Regarding individual PUFA contents, the linoleic acid (C18:2 n−6), linolenic acid (C18:3 n−3), EPA (C20:5 n−3), docosapentaenoic acid (C22:5 n−3), and DHA (C22:6 n−3) contents were greater in the ALF group than in the CON group ($P < 0.05$). In addition, the LT muscle of ALF lambs had a greater n−3 PUFA content and a lower n−6:n−3 ratio ($P < 0.001$) when compared with those in CON lambs but did not affect the total PUFA content ($P = 0.07$).

Microarray Gene Expression Results

Identification and Classification of Differentially Expressed Genes in LT Muscle, Liver, and SF

Forty-one, four and 96 genes were differentially expressed in LT muscle, SF, and liver, respectively, after SAM analysis (**Supplementary Figure 1**). In LT muscle, 41 genes were differentially expressed with an FDR = 0.002 (**Table 3**), of which

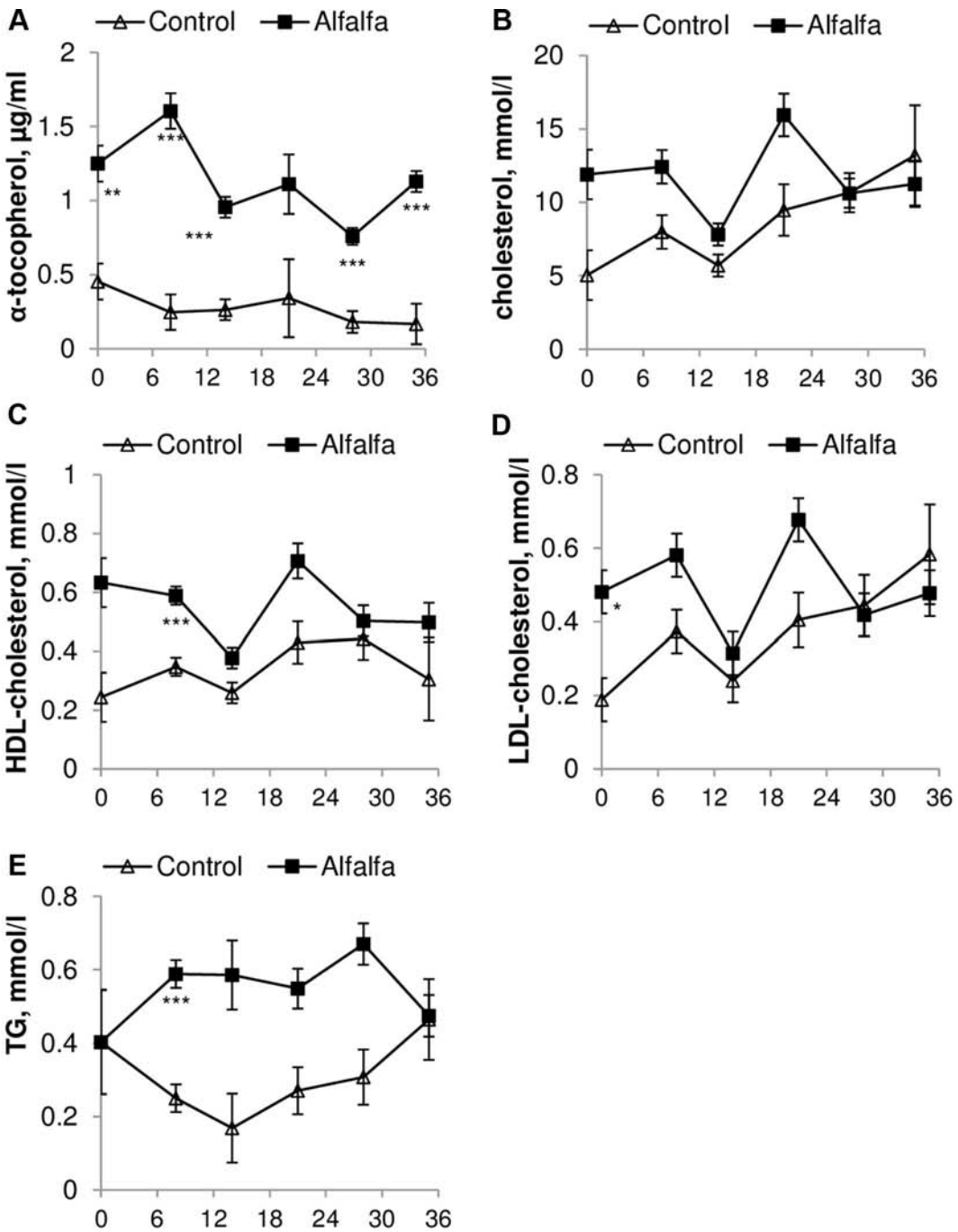


FIGURE 1 | Concentration of **(A)** α-tocopherol, **(B)** cholesterol, **(C)** HDL-cholesterol, **(D)** LDL-cholesterol and **(E)** TG during the entire experimental period compared with those in the CON group.

32 were downregulated and nine genes were upregulated. In the liver, 96 genes were differentially expressed (**Supplementary Table 5**), among which four genes were upregulated and 92 genes were downregulated with ALF treatment (FDR = 0.002). The top 20 significant genes in the liver are shown in **Table 4**.

Regarding SF, when ALF treatment was compared with CON, only four genes were differentially expressed with an FDR = 0.051, and all of them were upregulated in the ALF group (**Table 5**).

Treatment-Dependent Multivariate Analysis Results of Gene Expression in *Longissimus thoracis* Muscle, Liver and Subcutaneous Fat

Principal component (PC) analysis of the complete set showed that the first two PCs covered 81.1% of the observed variance of the sample set in LT muscle (**Figure 2A**). The clusters corresponding to gene expression profiles from ALF and CON groups were clearly separated from each other. Very similar

TABLE 2 | Effect of the treatment on the content of α-tocopherol and fatty acid (FA) composition of LT muscle in Rasa Aragonesa lambs.

Compounds[1]	Treatment[2]		P-value	
	CON	ALF	Feeding	Slaughter Age
α-tocopherol, mg/kg muscle	0.48 ± 0.17	2.38 ± 0.17	0.0001	0.38
C8:0	0.02 ± 0.008	0.04 ± 0.008	0.15	0.07
C10:0	0.17 ± 0.016	0.25 ± 0.016	0.0047	0.13
C12:0	0.42 ± 0.051	0.53 ± 0.051	0.16	0.43
C14:0	4.47 ± 0.318	5.17 ± 0.318	0.16	0.78
C16:0	24.97 ± 0.715	24.1 ± 0.715	0.42	0.41
C16:1	2.06 ± 0.127	1.83 ± 0.127	0.24	0.32
C17:0	0.96 ± 0.069	1.15 ± 0.069	*0.08*	0.48
C17:1	0.81 ± 0.047	0.82 ± 0.047	0.90	0.35
C18:0	11.76 ± 0.364	12.87 ± 0.364	*0.06*	0.01
C18:1 n–9	41.48 ± 0.947	36.65 ± 0.947	0.005	0.29
C18:1 n–7	1.01 ± 0.135	1.21 ± 0.135	0.33	0.47
C18:2 n–6tt	0.14 ± 0.047	0.53 ± 0.047	0.0001	0.24
C18:2 n–6ct	6.6 ± 0.517	6.65 ± 0.517	0.95	0.58
C18:3 n–6	0.07 ± 0.005	0.07 ± 0.005	0.90	0.27
C18:3 n–3	0.37 ± 0.24	2.24 ± 0.24	0.0002	0.93
C20:0	0.08 ± 0.01	0.13 ± 0.01	0.004	0.34
CLA	0.59 ± 0.065	0.85 ± 0.065	0.02	0.62
C20:1 n–9	0.11 ± 0.008	0.09 ± 0.008	0.21	0.08
C22:0	0.16 ± 0.021	0.19 ± 0.021	0.36	0.92
C20:4 n–6	2.67 ± 0.37	2.01 ± 0.37	0.24	0.47
C20:3 n–3	0.001 ± 0.009	0.02 ± 0.009	0.24	0.76
C20:5 n–3	0.13 ± 0.12	0.9 ± 0.12	0.001	0.92
C24:0	0.24 ± 0.056	0.15 ± 0.056	0.31	0.78
C22:5 n–3	0.51 ± 0.114	1.05 ± 0.114	0.008	0.80
C22:6 n–3	0.21 ± 0.072	0.51 ± 0.072	0.02	0.56
Saturated FA (SFA)	43.25 ± 0.924	44.58 ± 0.924	0.35	0.10
Monounsaturated FA	45.46 ± 0.939	40.6 ± 0.939	0.005	0.25
Polyunsaturated FA (PUFA)	11.29 ± 1.187	14.82 ± 1.187	0.07	0.67
PUFA n–6	9.48 ± 0.855	9.25 ± 0.855	0.86	0.57
PUFA n–3	1.22 ± 0.521	4.72 ± 0.521	0.0008	0.93
n–6:n–3	7.67 ± 0.697	2.84 ± 0.697	0.0006	0.16
PUFA : SFA	0.26 ± 0.032	0.34 ± 0.032	0.14	0.54

[1]Fatty acid composition it is expressed as weight percentages total fatty acid methyl esters (g/100 g of FAME).
[2]CON, weaned lambs fed commercial concentrates; ALF, unweaned grazing alfalfa lambs.

results were obtained in the liver (**Figure 2B**), but in SF, this separation was less clear (**Figure 2C**).

Hierarchical Clustering Analysis (HCA) in *Longissimus Thoracis* Muscle and Liver

Hierarchical clustering analysis for gene expression was performed using all genes and only the significant genes for each tissue. Because only four genes were significant in SF, the results of cluster analysis are not included. The results of HCA using only the significant genes for LT muscle and liver are presented in **Figure 3**. The expression profile of these genes was able to cluster and correctly classify the samples within their corresponding group. The heatmap shows the presence of two different clusters in both tissues. These two clusters clearly distinguished the ALF group from the CON group, as both groups showed very different gene expression patterns. For example, in LT muscle, the genes *BOLA*, *HSF2*, *CHP1*, *DNAJB11*, *CDC5L*, *TP53INP2*, *CPT1B*, *C8ORF4*, and *NMT1* were upregulated in the ALF group. Furthermore, a second cluster including the rest of the genes was found to be downregulated in the ALF group (**Figure 3A**).

In the liver, the genes *BHMT*, *LOC105614373* and *SLC19A1* were upregulated in the ALF group, and a second cluster, including the remaining genes, was downregulated in the ALF group (**Figure 3B**).

Functional Clustering Annotation

Longissimus thoracis Muscle

To gain insight into the biological processes that are regulated differentially between dietary treatments, we performed enrichment analyses using DAVID and ClueGo. The results of DAVID functional annotation clustering (FAC) revealed that the most enriched functional clusters were associated with "lipid and catabolic processes" (*CPT1B*, *PLA2G16*, *SPSB1*, *LRTOMT*, *PLCD4*, *FBXO9*, *CNBP* and *CYP27A1*) and 14 genes related to "muscle development" (*ALDH2*, *ANK3*, *CPT1B*, *FZD7*, *HSF2*, *IGF1R*, *LRTOMT*, *MSTN*, *MYLK2*, *MYOD1*, *MYOZ1*, *NMT1*, *PRDM1* and *RSC1*) (**Supplementary Table 5**). All these genes were downregulated in the ALF group except *CPT1B*, *HSF2* and *NMT1*, although the confident enrichment scores were less than 1.3 in both cases. The biological roles of downregulated genes

TABLE 3 | Differentially expressed genes in *Longissimus thoracis* muscle and fold-change in ALF–CON[1] contrast.

Gene symbol	Gene name	q value[2]	FC[3]
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	0.001	–1.42
MYOD1	Myogenic differentiation 1	0.001	–2.77
NEK7	NIMA (never in mitosis gene a)-related kinase 7	0.001	–1.28
CDC5L	CDC5 cell division cycle 5-like	0.001	1.15
ANGEL1	Angel homolog 1	0.001	–1.18
BOLA3	Similar to bolA-like 3; bolA homolog 3	0.001	1.20
C8orf4	Chromosome 8 open reading frame 4	0.001	3.30
CHP1	Chromo domain-containing protein 1	0.001	1.26
CNBP	CCHC-type zinc finger, nucleic acid binding protein	0.001	–1.22
CPT1B	Choline kinase beta; carnitine palmitoyltransferase 1B (muscle)	0.001	1.47
DNAJB11	DnaJ (Hsp40) homolog, subfamily B, member 11	0.001	1.34
FZD7	Frizzled homolog 7	0.001	–1.33
HSF2	Heat shock transcription factor 2	0.001	1.47
IGF1R	Insulin-like growth factor 1 receptor	0.001	–1.19
LRTOMT	Leucine rich transmembrane and 0-methyltransferase domain containing	0.001	–1.23
MACROD1	MACRO domain containing 1	0.001	–1.17
MYLK2	Myosin light chain kinase 2	0.001	–1.39
NMT1	N-myristoyltransferase 1	0.001	1.11
PLA2G16	Phospholipase A2, group XVI	0.001	–1.36
PLEKHH3	Pleckstrin homology domain containing, family H member 3	0.001	–1.60
PRDM1	PR domain containing 1, with ZNF domain	0.001	–1.52
R3HCC1	R3H domain and coiled-coil containing 1	0.001	–1.18
RSC1A1	Regulatory solute carrier protein, family 1, member 1	0.001	–1.25
SPSB1	splA/ryanodine receptor domain and SOCS box containing 1	0.001	–1.39
TP53INP2	Tumor protein p53 inducible nuclear protein 2	0.001	1.66
ANK3	Ankyrin 3, node of Ranvier (ankyrin G)	0.001	–1.25
BCL9L	B-cell CLL/lymphoma 9-like	0.001	–1.33
DNPEP	Aspartyl aminopeptidase	0.001	–1.24
TCEA3	Transcription elongation factor A (SII), 3	0.001	–1.27
SLC7A8	Solute carrier family 7, member 8	0.001	–1.98
YPEL2	Yippee-like 2 (Drosophila)	0.001	–1.31
SLC8A3	Solute carrier family 8 (sodium/calcium) exchanger), member 3	0.001	–1.91
PLCD4	Phospholipase C, delta 4	0.001	–1.38
CMBL	Carboxymethylenebutenolidase homolog	0.001	–1.44
MYOZ1	Myozenin 1	0.001	–1.13
MLF2	Myeloid leukemia factor 2	0.001	–1.17
SF3A1	Splicing factor 3a, subunit 1, 120 kDa	0.001	–1.17
ALDH2	Aldehyde dehydrogenase 2 family (mitochondrial)	0.001	–1.21
CYP27A1	Cytochrome P450, family 27, subfamily A, polypeptide 1	0.002	–1.26
FBXO9	F-box protein 9	0.002	–1.18
MSTN	Myostatin	0.002	–2.43

[1]*CON, weaned lambs fed commercial concentrates; ALF, unweaned grazing alfalfa lambs.*
[2]*q value: significance level.*
[3]*FC, fold-change.*

in LT muscle were also visualized with ClueGO (**Figure 4**). The size of the nodes reflects the statistical significance of the term. The most enriched biological process was that of "skeletal muscle tissue development" with four genes, *MSTN*, *MYLK2*, *MYOD1* and *BCL9L*.

TABLE 4 | Top 20 differentially expressed genes in the liver and fold-change in ALF–CON[1] contrast.

Gene symbol	Gene name	q value[2]	FC[3]
ACACA	acetyl-CoA carboxylase alpha	0.001	–1.99
ACSS2	acyl-CoA synthetase short-chain family member 2	0.0007	–2.51
CSAD	cysteine sulfinic acid decarboxylase	0.0007	–3.69
FADS1	fatty acid desaturase 1	0.001	–1.81
FADS2	fatty acid desaturase 2	0.0006	–1.84
GLEAN_10000268	GLEAN_10000268	0.001	–1.91
GUCA2B	guanylate cyclase activator 2B (uroguanylin)	0.001	–2.11
HMGCS1	3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble)	0.001	–2.21
IFI6	interferon, alpha-inducible protein 6	0.001	–2.27
KCTD14	potassium channel tetramerisation domain containing 14	0.0006	–1.82
LOC100125623	period 3	0.001	–2.09
LOC105614373	cytochrome P450 2B11-like	0.0006	1.57
MID1IP1	MID1 interacting protein 1	0.001	–2.49
MTHFD1L	monofunctional C1-tetrahydrofolate synthase, mitochondrial-like	0.0006	–1.63
OSGIN1	oxidative stress induced growth inhibitor 1	0.001	–2.15
PPP1R3B	protein phosphatase 1, regulatory subunit 3B	0.001	–2.3
SCD	stearoyl-CoA desaturase (delta-9-desaturase)	0.001	–4.44
SLC19A1	solute carrier family 19 (folate transporter), member 1	0.0007	1.4
THRSP	thyroid hormone responsive	0.0006	–6.86
ZNF564	zinc finger protein 564	0.001	–2.52

[1]*CON: weaned lambs fed commercial concentrates; ALF: unweaned grazing alfalfa lambs.*
[2]*q value: significance level.*
[3]*FC, fold-change.*

TABLE 5 | Significant differentially expressed genes in subcutaneous fat and fold change in ALF–CON[1] contrast.

Gene symbol	Gene Name	q value[2]	FC[3]
CLDN12	claudin 12	0.01	1.26
LOC101121796	cytochrome P450 2B19-like	0.047	1.70
METTL1	methyltransferase like 1	0.047	1.33
GPN2	GPN-loop GTPase 2	0.047	1.31

[1]*CON, weaned lambs fed commercial concentrates; ALF, unweaned grazing alfalfa lambs.*
[2]*q value: significance level.*
[3]*FC, fold-change.*

Liver

The results of DAVID revealed two major gene clusters associated with "sterol biosynthesis" (*EBP*, *MVD*, *HMGCR*, *CYP51A1*, *HMGCS1*, *NR0B2*, *C14ORF1*, *FDFT1*, *SQLE*, *DHCR7*, *SC5DL*, *DHCR24*, and *NSDHL*), "lipid biosynthetic process" (*ACACA*, *CYP51A1*, *FADS1*, *FADS2*, *SCD* and *SC5DL*), and "cholesterol

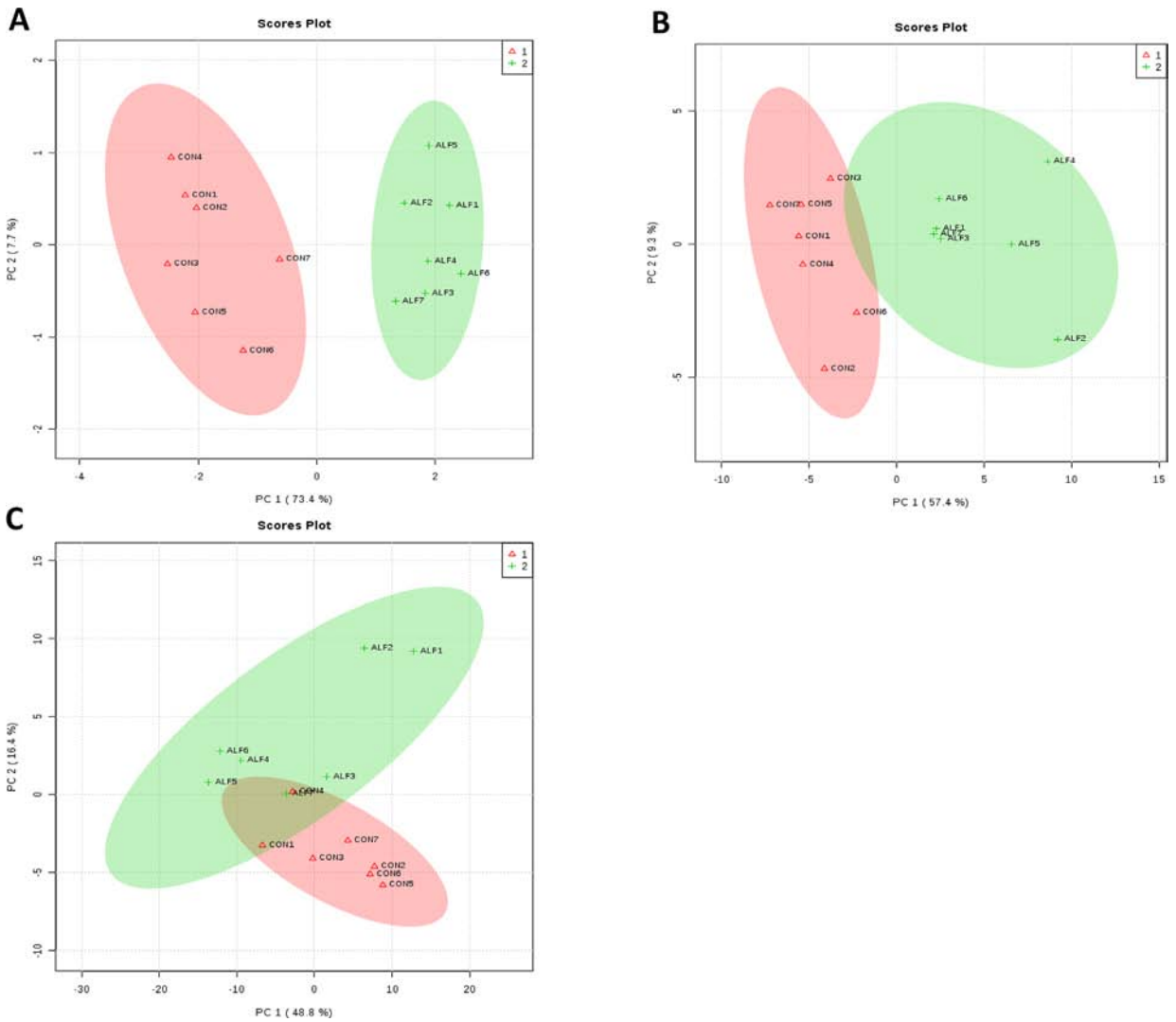


FIGURE 2 | Principal components analysis (PCA) based on gene expression profile data for the **(A)** LT muscle **(B)** liver, and **(C)** subcutaneous fat. Principal components analysis scores plots discriminating between the LT muscle of lambs fed concentrate CON (+) and Alfalfa (triangle).

metabolic process" (*EBP*, *MVD*, *HMGCR*, *CYP51A1*, *SQLE*, *DHCR7*, *HMGCS1*, *NR0B2*, *DHCR24*, *FDFT1*, and *NSDHL*) (**Supplementary Table 6**). Similar results were obtained with ClueGo, where the most enriched biological processes were "fatty acid biosynthetic process," "sterol metabolic process," "cofactor metabolic processes" and "coenzyme metabolic processes" (**Figure 5**). These genes were all downregulated in ALF treatment.

Subcutaneous Fat

Only 4 genes were significant in SF, and no cluster was found with DAVID FAC.

Validation of Microarray Results Using qPCR

The gene set selected to validate the microarray results by qPCR included the following 15 genes: *CPT1B*, *MYOD1*, *MSTN*, *ABCC4*, *IGF1R* and *PLA2G16* for LT muscle; *FADS1*, *FADS2*, *ACACA*, *SCD*, *SQLE*, *IER3*, *SLC19A1* and *THRSP* for liver; and *METTL1* for SF. The genes were selected because they were significantly differentially expressed between groups. The expression of these genes using microarray technology and qPCR is shown in **Table 6**. The housekeeping genes *GUSB* and *YWHAZ* were used to normalize the results for LT muscle and SF. In the liver, five candidate reference genes were tested, and the most stable genes exhibiting the lowest expression stability value (M) were *RPL37* (M = 0.182), *GUSB* (M = 0.226), and *RPL19* (M = 0.296). The three reference genes were more stable than the GOIs. The magnitude of the fold change obtained by microarray and qPCR was slightly different in some instances, but the qPCR results demonstrated a similar trend compared with the microarray results of these genes (**Table 6**).

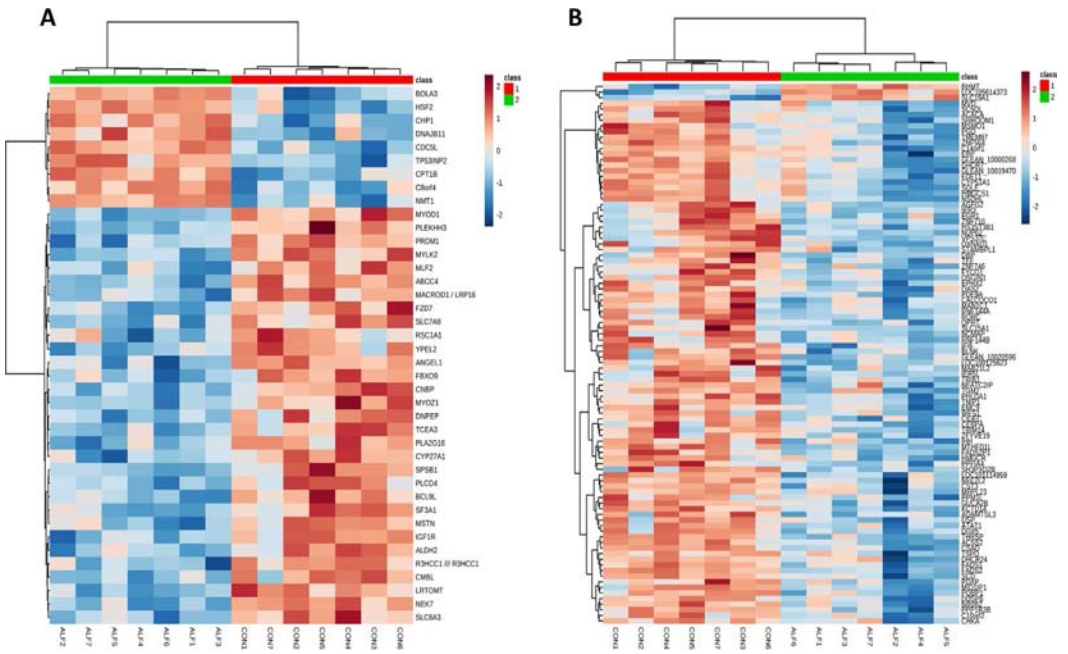


FIGURE 3 | Hierarchical clustering analysis of significant gene expression in the **(A)** LT muscle and **(B)** liver of CON lambs (1) and ALF lambs (2). Cells are colored based on the measured signal intensity. Dark brown represents high gene expression levels, blue indicates low signal intensity, and gray cells represent the intermediate level.

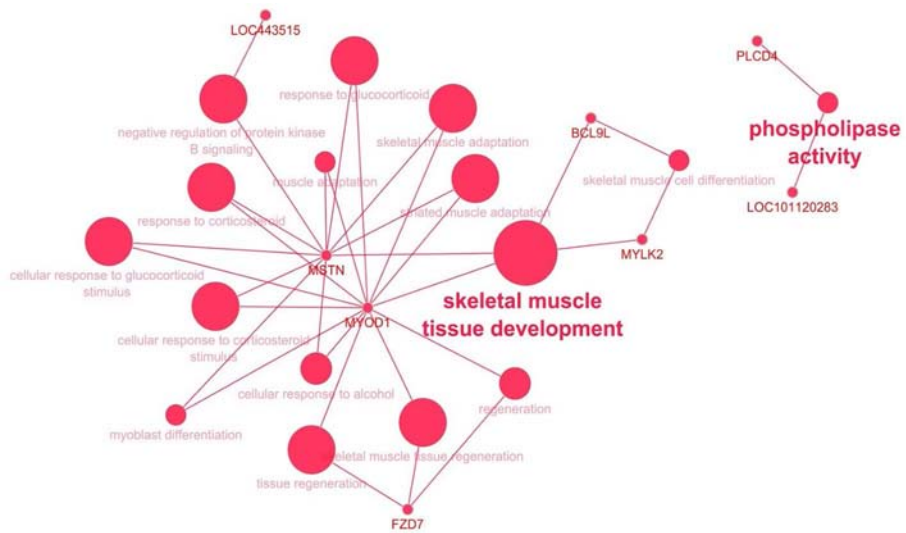


FIGURE 4 | The biological role of up- (green) and downregulated (red) genes in the LT muscle of ALF animals visualized with ClueGO. The size of the nodes reflects the statistical significance of the term.

DISCUSSION

In this study, we investigated the fatty acid profile and gene expression using a microarray in the LT muscle, liver and SF of lambs fed concentrate or alfalfa. As expected, ALF animals contained greater CLAs and a greater proportion of n−3 PUFAs in muscle, such as linolenic acid (C18:3 n−3), EPA (C20:5 n−3), docosapentaenoic acid (C22:5 n−3), and DHA (C22:6 n−3), when compared with levels in the CON group. Many studies have reported the impact of grazing on the fatty acid profile in meat lambs, particularly the fatty acids of the n−3 series (Fisher et al., 2000; Dervishi et al., 2010; Vasta et al., 2012). Zhang et al. (2017) suggested that specific compounds in the diet can be transferred to the meat. In our experiment, the fatty acid composition, especially that of the n−3 series, and α-tocopherol are probably a reflection of diet composition. Suckling lambs are functionally non-ruminants,

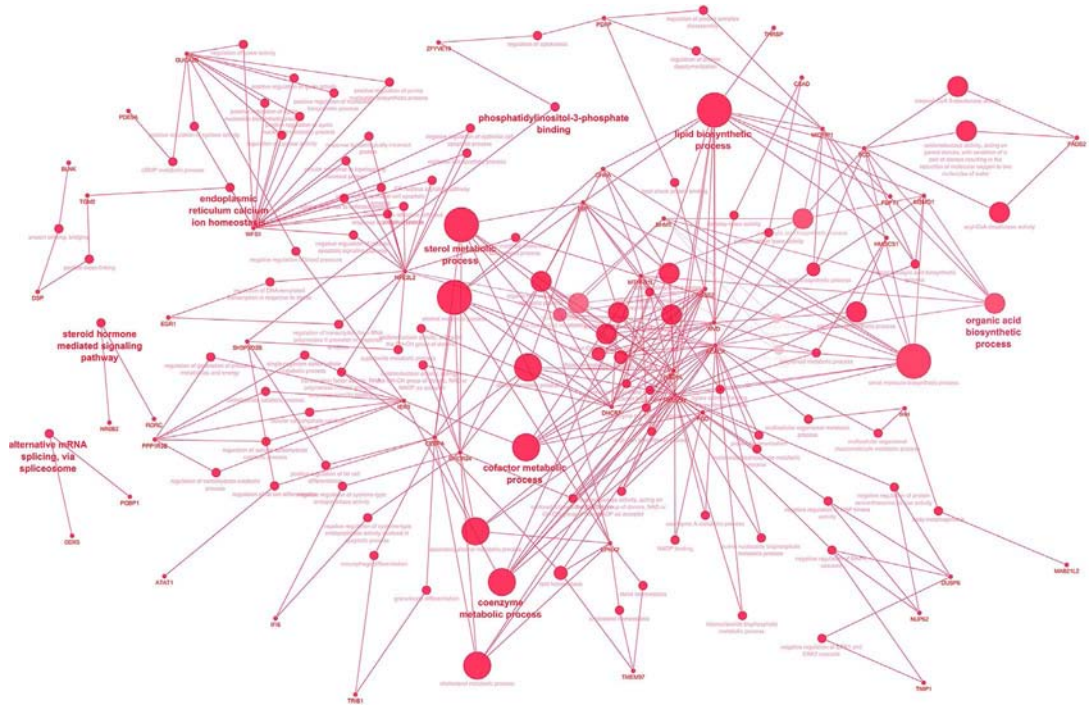


FIGURE 5 | The biological role of up- (green) and downregulated (red) genes in ALF animals visualized with ClueGO in the liver. The size of the nodes reflects the statistical significance of the term.

TABLE 6 | Real-time PCR confirmation of the microarray results. Gene expression changes in LT muscle, liver and subcutaneous fat in ALF vs CON comparison, and the fold change (FC) obtained with microarray and qPCR data.

Genes	Microarray	qPCR
LT muscle		
CPT1B	1.47**	1.15
MYOD1	–2.77**	–2.03*
MSTN	–2.43**	–5.73**
ABCC4	–1.42**	–1.89†
IGF1R	–1.19**	–1.79
PLA2G16	–1.36**	–2.27**
Subcutaneous fat		
METTL1	1.33*	1.55
Liver		
FADS1	–1.81**	–1.65
FADS2	–1.82**	–2.31*
ACACA	–1.99**	–1.52
SCD	–4.43**	–3.78*
SQLE	–1.42**	–1.89†
IER3	–1.95**	–1.45
SLC19A1	1.41**	1.21
THRSP	-6.86**	–6.13**

*$P < 0.05$, ** $P < 0.01$ and †$0.05 < P < 0.10$.

and their meat FA profile should reflect the FA profile of the suckled milk (Napolitano et al., 2002; Valvo et al., 2005). Thus, grazing dams is an advisable alternative to increase PUFAs in the suckling lamb meat because fresh pasture has a high concentration of linolenic acid (C18:3n–3), which increases the contents of vaccenic acid (C18:1t-11), conjugated linoleic acid isomers (CLA), and n–3 PUFA in milk compared with diets comprising concentrate or preserved forage (Nudda et al., 2005; Joy et al., 2012). The high value of C18:3n–3 in ALF lamb meat could be due to the C18:3n–3 provided by pasture that, as they are not yet fully weaned, is not bio-hydrogenated by rumen microbiota. Moreover, the relatively low effectiveness of milk in affecting meat fatty acid composition, could explain the slightly difference between CON and ALF in CLA and VA. Therefore, lambs that were allowed to graze resulted in a meat fatty acid profile that is richer in fatty acids of the n–3 series, mainly due to the dam's milk that were grazing continuously during lactation in alfalfa pastures. According to Álvarez-Rodriguez et al. (2018) dietary alfalfa but not milk supply improved CLA, and n–3 PUFAs contents in lamb meat. The FA composition of ALF lambs was more related to ewe's milk than to fresh forage (Dervishi et al., 2010). Previous studies have shown that grazing increases the PUFA content in milk, particularly linolenic acid (C18:3n–3), while concentrates modify rumen retention time of the feed, increase linoleic acid (C18:2n–6) intake, and alter biohydrogenation pathways toward lower n–3 PUFA and CLA contents, leading to lower contents of these compounds in the milk (Elgersma, 2015). In addition, these animals had greater α-tocopherol in muscle and plasma. Vitamin E is a powerful fat-soluble antioxidant that plays important roles in scavenging free radicals and neurologic function (Wang and Quinn, 2000; Traber and Atkinson, 2007). In this study, we found that lipid oxidation was lower in ALF lambs on day 7 of display when compared with the levels in CON lambs. These results are in concordance with previous studies in which we reported that the addition of vitamin E to the diet increased the

α-tocopherol muscle content and drastically diminished the lipid oxidation of meat (Kasapidou et al., 2012; González-Calvo et al., 2015b; Ponnampalam et al., 2017).

Moreover, we investigated how the feeding system impacted gene expression in LT muscle, liver and SF in both treatment groups. Indeed, we found that both groups differed in their gene expression profile, mainly in LT muscle and liver, with the greatest impact in liver. It has been reported that dietary intervention can lead to major changes in gene expression in muscle and liver (Dervishi et al., 2011; Cui et al, 2018). In the ALF group, the most enriched biological processes in LT muscle were skeletal muscle tissue development (*MYOD1*, *MYLK2* and *MSTN*) (**Figure 4** and **Supplementary Table 5**). These genes were downregulated in the ALF group, with *MYOD1* and *MSTN* being the most downregulated genes (the lowest FC). The yield of saleable meat and meat quality, and therefore the profitability for livestock operations, are greatly influenced by growth during the postnatal period. Therefore, the identification of genes that play a role in muscle growth in sheep is an important step for improving sheep meat production by selection. In this regard, in livestock species, *MYOD1* and *MSTN* are considered candidate genes for meat quality and carcass traits (Ibeagha-Awemu et al., 2008; Bhuiyan et al., 2009). *MYOD1* regulates muscle cell differentiation, growth, and development and is also involved in muscle regeneration (Kitzmann et al., 1998). For example, polymorphisms of *MYOD1* have been associated with weight, several muscle fiber characteristics, the loin eye area and lightness in yak populations, pork and cattle (Chu et al., 2012; Lee et al., 2012; Du et al., 2013). In addition, low *MYOD1* expression levels were related to low Warner–Bratzler shear force measured in the *longissimus dorsi* muscle of beef (Tizioto et al., 2014) and thus with greater tenderness. In sheep, a positive correlation between *MYOD1* expression and cold carcass yield was found (Lôbo et al., 2012). The authors proposed that animals with a higher expression of *MYOD1* were more efficient during postnatal growth and had a greater *longissimus* dorsi weight and a better cold and hot carcass yield. In our study, we did not observe differences in slaughter weight or average daily gain between the ALF and CON groups. These discrepancies may be due to different slaughter ages among both studies. We sacrificed our animals at 67–72 days, and other studies compared heavy lambs (at an average of 200 days) fed either concentrate or limited grazing (Lôbo et al., 2012), whereas in the present study, grazing lambs had free access to forage, concentrate, and dam's milk. Despite the different results, further investigation into the role of *MYOD1* in sheep carcass and meat quality traits in sheep is necessary for effective marker assisted selection. Another gene that was downregulated in the ALF group was *MSTN*. Myostatin is an extracellular cytokine that is mostly expressed in skeletal muscles and is known to play a crucial role in the negative regulation of muscle mass (Elkina et al., 2011). This effect is due to an increase in both muscle fiber number (hyperplasia) and mass (hypertrophy). For instance, Ji et al. (1998) demonstrated that myostatin expression in skeletal muscle peaks prenatally and that greater expression during the prenatal period is associated with low birth weight in pigs. Mutations in the myostatin gene with functional inactivation in beef cattle increase the muscle mass in the double-muscled phenotype and lead to smaller adipocytes and fewer fat islands in muscle (Cassar-Malek et al., 2007). In addition, in different cattle breeds, mutations in *MSTN* have been associated with significant reductions in the shear force and a decrease in total collagen content (Ngapo et al., 2002; Lines et al., 2009). Moreover, mutations in *MSTN* in sheep were associated with muscling and reduced intramuscular fat (Kijas et al., 2007) and an increased percentage of fast glycolytic myofibers (Laville et al., 2004). In our experiment, the ALF group showed downregulated *MYOD1* and *MSTN* genes, which may be beneficial for increasing meat tenderness and cold carcass yield in heavier animals. However, the simultaneous downregulation of the *MYOD1* and *MSTN* genes in ALF group might determine an opposite effect about animal's performance, thus justifying the lack in different performance between two groups.

The results of the functional analysis showed that genes related to catabolic and lipid processes in LT muscle were downregulated (*PLA2G16*, *SPSB1*, *LRTOMT*, *PLCD4*, *FBXO9*, *CNBP* and *CYP27A1*), except for *CPT1B*, which was upregulated in the ALF group (**Supplementary Table 5**). Carnitine palmitoyl transferase I (M-CPT 1), codified by the *CPT1B* gene, is part of the mitochondrial transport system and is a key enzyme in the control of long-chain fatty acid oxidation (Bartelds et al., 2004). These results are in agreement with those previously obtained by Dervishi et al. (2011, 2012) in which grazing systems promoted higher levels of *CPT1B* gene expression in the *semitendinosus* muscle and mammary gland. As reported by Dervishi et al. (2010), concentrate feeding promotes the upregulation of genes related to adipogenesis, whereas the grazing system promotes higher levels of genes implicated in fatty acid oxidation.

The impact of feeding system was more pronounced on liver gene expression, where 96 genes were significantly changed compared to that in LT muscle (41 genes) and SF (4 genes). The major sites of fatty acid synthesis are adipose tissue and the liver. However, the results for gene expression in these three different tissues suggest that in young lambs, the major site of lipid metabolism is the liver rather than subcutaneous fat.

We attempted to link the significant fatty acids in muscle and metabolites in plasma with the results of gene expression to obtain a better understanding of the underlying metabolic processes associated with different feeding systems. The relationship between n−3 FAs in muscle and gene expression in liver mapped 30 significant genes (out of 92) and 9 significant compounds (out of 13) to KEGG IDs. A compound-gene network was generated (**Figure 6** and **Supplementary Figure 2**). In addition, this approach helped us to identify genes related to enriched biological processes and certain desired outcomes, for example, n−3 PUFA series that are desirable regarding human health (Simopoulos, 2008; Cabo et al., 2012; Liu and Ma, 2014). Indeed, the animals grazing alfalfa had a greater content of fatty acids of the n−3 series such as linolenic acid (C18:3 n−3), EPA (C20:5 n−3), docosapentaenoic acid (C22:5 n−3), and DHA (C22:6 n−3) in LT muscle and lower expression of *FADS1* and *FADS2* in liver (**Figure 6**). It is worth mentioning that *FADS1* and *FADS2* in the "fatty acid biosynthetic process" cluster are key genes in the metabolism of n−3 PUFA series. The proteins

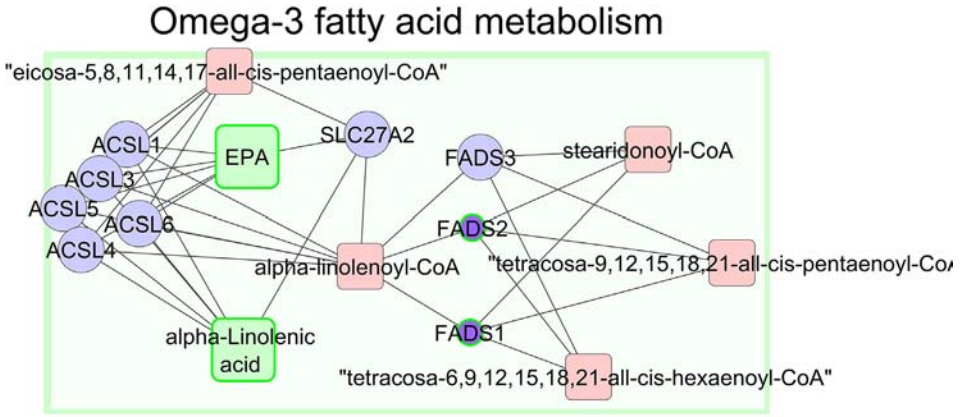


FIGURE 6 | The network of metabolites and genes involved in omega-3 fatty acid metabolism in ALF lambs. Significant metabolites with experimental data are shown in green squares, and significant genes with experimental data are shown in purple circles with green borders. The size of the nodes represents the direction of the change. A small purple circle with a green border indicates downregulated genes, and large green square nodes point to upregulated compounds.

encoded by these genes (*FADS1* and *FADS2*) are members of the fatty acid desaturase (FADS) gene family. Desaturase enzymes regulate the unsaturation of fatty acids through the introduction of double bonds between defined carbons of the fatty acyl chain (Nakamura and Nara, 2004). In addition, these animals were characterized by a decrease in the expression of genes related to cholesterol metabolism (*DHCR7*, *SC5DL*, *EBP*, *NSDHL*, *MTHFD1L*, and *CYP51A1*; **Supplementary Figure 2**).

Nutrition is an important strategy to alter gene expression and the fatty acid profile of meat. It has been widely reported that grazing animals have a greater content of the n−3 PUFA series in the serum, liver and muscle and a lower n−6:n−3 ratio, in agreement with the present study. Interestingly, we also found that the expression of two genes related to n−3 PUFA metabolism was downregulated in the livers of ALF animals. *Fatty acid desaturase 1* (*FADS1*) and *2* (*FADS2*) genes encode delta-5 and delta-6 desaturases, respectively, which are rate-limiting enzymes in the synthesis of polyunsaturated omega-3 and omega-6 FAs. Dietary FAs have been shown to regulate desaturase activity (Nakamura and Nara, 2004). Gene expression of both *FADS1* and *FADS2* is reduced by PUFAs in several hepatic models (Reardon et al., 2013; Cho et al., 1999a; Cho et al., 1999b). Furthermore, *FADS1* and *FADS2* gene expression was reduced by EPA and AA in 3T3-L1 adipocytes (Ralston et al., 2015). In our study we found that ALF lambs have greater amount of EPA in their muscle mainly because of their diet. ALF lambs ingested diets rich in PUFAs (fresh alfalfa, and mainly dams' milk), which in turn might have down-regulated *FADS1* and *FADS2* gene expression in liver. In support to our speculation da Costa et al. (2014) found that high levels of n−3 PUFA in cattle liver down-regulated the expression of the genes *FADS1* and *FADS2*.

The results found in the current study showed that ingesting diets richer in n−3 PUFA might have negative effects on the *de novo* synthesis of n−3 PUFA by the FADS1 and FADS2 enzymes. However, feeding diets poorer in n−3 PUFA can promote fatty acid desaturation, which makes these two genes attractive candidates for altering the content of PUFAs in meat, by looking for polymorphisms that may affect the functionality and efficiency of these enzymes and alter the fatty acid profile in lamb meat. Functional SNPs can provide an additional resource as a potential genetic markers in breeding programs. In this respect, in humans, numerous studies have consistently replicated the associations between polymorphisms in the *FADS1* and *FADS2* genes and the PUFA concentration (Corella and Ordovas, 2012). In porcine, a polymorphism in exon 3 of the pig *FADS2* has been associated with C20:4 and intramuscular fat (IMF) content (Renaville et al., 2013; Gol et al., 2018). In dairy cows, Ibeagha-Awemu et al. (2014) demonstrated positive associations between three SNPs within *FADS1* and *FADS2* with three milk PUFAs. Meanwhile, contradictory results have been reported in sheep. For example, a SNP in *FADS2* was significantly associated with intramuscular levels of EPA (C20:5n−3) and DHA (C22:6n−3) (Malau-Aduli et al., 2011), but in a different report, no SNP within the *FADS1* and *FADS2* gene regions was associated with lamb muscle n−3 levels (Knight et al., 2012). Our study further points to the importance of nutritional modulation of *FADS1* and *FADS2* gene expression and the fatty acid profile in sheep.

CONCLUSION

Grazing lambs presented a higher content of CLA and n−3 PUFA series and showed a lower n−6/n−3 ratio, which is favorable with regard to current human health. The feeding system is the main factor affecting the fatty acid composition and gene expression in LT muscle and liver. The gene expression results in the three different tissues suggest that the major site of lipid metabolism is the liver rather than subcutaneous fat in young lambs of the Rasa Aragonesa breed. Gene expression of *FADS1* and *FADS2* plays an important role in the synthesis of n−3 PUFA series, which in turn makes these two genes attractive candidates to alter the content of PUFAs in meat. More studies will be necessary to elucidate the effects of the feeding system on *FADS1* and *FADS2* expression in other tissues of interest or to search for mutations or functional SNPs that may be used in the future as a tool to improve the fatty acid profile in lamb meat.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in NCBI Gene Expression Omnibus repository, GSE63774 (LT muscle and SF) and GSE125661 (liver).

ETHICS STATEMENT

All Experimental Procedures, Including the Care of Animals and Euthanasia, Were Performed in Accordance With the Guidelines of the European Union and Spanish Regulations for the Use and Care of Animals in Research and Were Approved by the Animal Welfare Committee of the Centro De Investigación Y Tecnología Agroalimentaria (CITA) (Protocol Number 2009-01_MJT). in All Cases, Euthanasia Was Performed by Penetrating Captive Bolt Followed by Immediate Exsanguination.

AUTHOR CONTRIBUTIONS

LG-C, PS, and MB performed the experiments. MJ, MS, and JC designed the research and obtained funding for this research. ED, MS, MB, and JC wrote the paper. ED, LG-C, MS, MB, MJ, RM-H, JO, and JC analyzed the data. MJ provided animals. JC had primary responsibility for the final content. All of the authors contributed to the manuscript discussion. All of the authors read and approved the final manuscript.

FUNDING

This study was supported by the Ministry of Economy and Competitiveness of Spain, the European Union Regional Development Funds (INIA RTA2012-080-00, INIA RZP2017-00001-00) and the Research Group Funds of Aragón Government (A14_17R).

ACKNOWLEDGMENTS

Appreciation is expressed to the staff of CITA de Aragón for their help in data collection. Special thanks are extended to F. Molino and G. Ripoll for their laboratory assistance

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01070/full#supplementary-material

REFERENCES

Álvarez-Rodríguez, J., Ripoll, G., Lobón, S., Sanz, A., Blanco, M., and Joy, M. (2018). Alfalfa but not milk in lamb's diet improves meat fatty acid profile and α-tocopherol content. *Food Res. Int.* 107, 708–716. doi: 10.1016/j.foodres.2018.03.007

Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 5245–5250. doi: 10.1158/0008-5472.CAN-04-0496

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). NCBI GEO: archive for functional genomics data sets update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193

Bartelds, B., Takens, J., Smid, G. B., Zammit, V. A., Prip-Buus, C., Kuipers, J. R. G., et al. (2004). Myocardial carnitine palmitoyltransferase I expression and long-chain fatty acid oxidation in fetal and newborn lambs. *Am. J. Physiol. Circ. Physiol.* 286, H2243–H2248. doi: 10.1152/ajpheart.00864.2003

Bhuiyan, M. S. A., Kim, N. K., Cho, Y. M., Yoon, D., Kim, K. S., Jeon, J. T., et al. (2009). Identification of SNPs in MYOD gene family and their associations with carcass traits in cattle. *Livest. Sci.* 126, 292–297. doi: 10.1016/J.LIVSCI.2009.05.019

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi: 10.1093/bioinformatics/btp101

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917. doi: 10.1139/o59-099

Cassar-Malek, I., Passelaigue, F., Bernard, C., Léger, J., and Hocquette, J.-F. (2007). Target genes of myostatin loss-of-function in muscles of late bovine fetuses. *BMC Genomics* 8, 63. doi: 10.1186/1471-2164-8-63

Cabo, J., Alonso, R., and Mata, P. (2012). Omega-3 fatty acids and blood pressure. *Br. J. Nutr.* 107, S195–S200. doi: 10.1017/S0007114512001584

Cho, H. P., Nakamura, M., and Clarke, S. D. (1999a). Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. *J. Biol. Chem.* 274, 37335–37339. doi: 10.1074/jbc.274.52.37335

Cho, H. P., Nakamura, M. T., and Clarke, S. D. (1999b). Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. *J. Biol. Chem.* 274, 471–477. doi: 10.1074/jbc.274.1.471

Chu, M., Yan, P., Liang, C. N., Guo, X., Pei, J., Ding, X. Z., et al. (2012). Association between polymorphism of MyoD1 gene and growth traits in Yak. *China Anim. Husb. Vet. Med* 29, 111–113.

Corella, D., and Ordovás, J. M. (2012). Interactions between dietary n-3 fatty acids and genetic variants and risk of disease. *Br. J. Nutr.* 107 Suppl 2(0 2), S271–S283. doi: 10.1017/S0007114512001651

Cui, K., Wang, B., Ma, T., Si, B. W., Zhang, N. F., Tu, Y., et al. (2018). Effects of dietary protein restriction followed by realimentation on growth performance and liver transcriptome alterations of lamb. *Sci. Reports.* 8, 15185. doi: 10.1038/s41598-018-33407-w

da Costa, A. S., Bessa, R. J., Pires, V. M., Rolo, E. A., Pinto, R. M., Andrade Fontes, C. M., et al. (2014). Is hepatic lipid metabolism of beef cattle influenced by breed and dietary silage level? *BMC Vet. Res.* 10, 65. doi: 10.1186/1746-6148-10-65

Dervishi, E., Joy, M., Alvarez-Rodriguez, J., Serrano, M., and Calvo, J. H. (2012). The forage type (grazing versus hay pasture) fed to ewes and the lamb sex affect fatty acid profile and lipogenic gene expression in the longissimus muscle of suckling lambs1. *J. Anim. Sci.* 90, 54–66. doi: 10.2527/jas.2011-4057

Dervishi, E., Serrano, C., Joy, M., Serrano, M., Rodellar, C., and Calvo, J. H. (2010). Effect of the feeding system on the fatty acid composition, expression of the Δ9-desaturase, peroxisome proliferator-activated receptor alpha, gamma, and sterol regulatory element binding protein 1 genes in the semitendinous muscle of light lambs of the R. *BMC Vet. Res.* 6, 40. doi: 10.1186/1746-6148-6-40

Dervishi, E., Serrano, C., Joy, M., Serrano, M., Rodellar, C., and Calvo, J. H. (2011). The effect of feeding system in the expression of genes related with fat metabolism in semitendinous muscle in sheep. *Meat Sci.* 89, 91–97. doi: 10.1016/j.meatsci.2011.04.003

Du, X. H., Gan, Q. F., Yuan, Z. R., Gao, X., Zhang, L. P., Gao, H. J., et al. (2013). Polymorphism of MyoD1 and Myf6 genes and associations with carcass and meat quality traits in beef cattle. *Genet. Mol. Res.* 12, 6708–6717. doi: 10.4238/2013.December.13.4

Elgersma, A. (2015). Grazing increases the unsaturated fatty acid concentration of milk from grass-fed cows: a review of the contributing factors, challenges

and future perspectives. *Eur. J. Lipid Sci. Technol.* 117, 1345–1369. doi: 10.1002/ejlt.201400469

Elkina, Y., von Haehling, S., Anker, S. D., and Springer, J. (2011). The role of myostatin in muscle wasting: an overview. *J. Cachexia. Sarcopenia Muscle* 2, 143–151. doi: 10.1007/s13539-011-0035-5

Fisher, A. V., Enser, M., Richardson, R. I., Wood, J. D., Nute, G. R., Kurt, E., et al. (2000). Fatty acid composition and eating quality of lamb types derived from four diverse breed × production systems. *Meat Sci.* 55, 141–147.

Gol, S., Pena, R. N., Rothschild, M. F., Tor, M., and Estany, J. (2018). A polymorphism in the fatty acid desaturase-2 gene is associated with the arachidonic acid metabolism in pigs. *Sci. Rep*. 8. doi: 10.1038/s41598-018-32710-w

González-Calvo, L., Dervishi, E., Joy, M., Sarto, P., Martin-Hernandez, R., Serrano, M., et al. (2017). Genome-wide expression profiling in muscle and subcutaneous fat of lambs in response to the intake of concentrate supplemented with vitamin E. *BMC Genomics* 18, 92. doi: 10.1186/s12864-016-3405-8

González-Calvo, L., Joy, M., Alberti, C., Ripoll, G., Molino, F., Serrano, M., et al. (2014). Effect of finishing period length with α-tocopherol supplementation on the expression of vitamin E-related genes in the muscle and subcutaneous fat of light lambs. *Gene* 552, 225–233. doi: 10.1016/j.gene.2014.09.037

González-Calvo, L., Joy, M., Blanco, M., Dervishi, E., Molino, F., Sarto, P., et al. (2015a). Effect of vitamin E supplementation or alfalfa grazing on fatty acid composition and expression of genes related to lipid metabolism in lambs1. *J. Anim. Sci.* 93, 3044–3054. doi: 10.2527/jas.2014-8758

González-Calvo, L., Ripoll, G., Molino, F., Calvo, J. H., and Joy, M. (2015b). The relationship between muscle α -tocopherol concentration and meat oxidation in light lambs fed vitamin E supplements prior to slaughter. *J. Sci. Food Agric.* 95, 103–110. doi: 10.1002/jsfa.6688

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Hunt, M. (1980). Meat color measurements. *Proceedings of 33th Reciprocal Meat Conference* (1980). *Reciprocal Meat Conf. Proc.* 33, 41–46. Avalaible at: https://meatscience.org/docs/default-source/publications-resources/rmc/1980/meat-color-measurements.pdf?sfvrsn=ae05bbb3_2. [Accessed October 18, 2019].

Ibeagha-Awemu, E. M., Akwanji, K. A., Beaudoin, F., and Zhao, X. (2014). Associations between variants of FADS genes and omega-3 and omega-6 milk fatty acids of Canadian Holstein cows. *BMC Genet.* 15, 25. doi: 10.1186/1471-2156-15-25

Ibeagha-Awemu, E. M., Kgwatalala, P., and Zhao, X. (2008). A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm. Genome* 19, 591–617. doi: 10.1007/s00335-008-9141-x

Joy, M., Ripoll, G., Molino, F., Dervishi, E., and Álvarez-Rodriguez, J. (2012). Influence of the type of forage supplied to ewes in pre- and post-partum periods on the meat fatty acids of suckling lambs. *Meat Sci.* 90, 775–782. doi: 10.1016/J.MEATSCI.2011.11.013

Ji, S., Losinski, R. L., Cornelius, S. G., Frank, G. R., Willis, G. M., Gerrard, D. E., et al. (1998). Myostatin expression in porcine tissues: tissue specificity and developmental and postnatal regulation. *Am. J. Physiol. Integr. Comp. Physiol.* 275, R1265–R1273. doi: 10.1152/ajpregu.1998.275.4.R1265

Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., et al. (2012). Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* 28, 373–380. doi: 10.1093/bioinformatics/btr661

Kasapidou, E., Wood, J. D., Richardson, R. I., Sinclair, L. A., Wilkinson, R. G., and Enser, M. (2012). Effect of vitamin E supplementation and diet on fatty acid composition and on meat colour and lipid oxidation of lamb leg steaks displayed in modified atmosphere packs. *Meat Sci*. 90, 908–916. doi: 10.1016/j.meatsci.2011.11.031

Kijas, J. W., McCulloch, R., Edwards, J. E. H., Oddy, V. H., Lee, S. H., and van der Werf, J. (2007). Evidence for multiple alleles effecting muscling and fatness at the Ovine GDF8 locus. *BMC Genet.* 8, 80. doi: 10.1186/1471-2156-8-80

Kitessa, S., Liu, S., Briegel, J., Pethick, D., Gardner, G., Ferguson, M., et al. (2010). Effects of intensive or pasture finishing in spring and linseed supplementation in autumn on the omega-3 content of lamb meat and its carcass distribution. *Anim. Prod. Sci.* 50, 130. doi: 10.1071/AN09095

Kitzmann, M., Carnac, G., Vandromme, M., Primig, M., Lamb, N. J., and Fernandez, A. (1998). The muscle regulatory factors MyoD and myf-5 undergo distinct cell cycle-specific expression in muscle cells. *J. Cell Biol.* 142, 1447–1459.

Knight, M. I., Daetwyler, H. D., Hayes, B. J., Hayden, M. J., Ball, A. J., Pethick, D. W., et al. (2012). Discovery and trait association of single nucleotide polymorphisms from gene regions of influence on meat tenderness and long-chain omega-3 fatty acid content in Australian lamb. *Anim. Prod. Sci.* 52, 591. doi: 10.1071/AN11229

Laville, E., Bouix, J., Sayd, T., Elsen, J. M., Larzul, C., et al. (2004). Effects of a quantitative trait locus for muscle hypertrophy from Belgian Texel sheep on carcass conformation and muscularity. *J. Anim. Sci.* 82, 3128–3137. doi: 10.2527/2004.82113128x

Lee, E. A., Kim, J. M., Lim, K. S., Ryu, Y. C., Jeon, W. M., and Hong, K. C. (2012). Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. *Meat Sci.* 92, 36–43. doi: 10.1016/J.MEATSCI.2012.03.018

Lehnen, T. E., da Silva, M. R., Camacho, A., Marcadenti, A., and Lehnen, A. M. (2015). A review on effects of conjugated linoleic fatty acid (CLA) upon body composition and energetic metabolism. *J. Int. Soc. Sports Nutr.* 12, 36. doi: 10.1186/s12970-015-0097-4

Li, K., Sinclair, A. J., Zhao, F., and Li, D. (2018). Uncommon fatty acids and cardiometabolic health. *Nutrients* 10 (10), 1559. doi: 10.3390/nu10101559

Lines, D. S., Pitchford, W. S., Kruk, Z. A., and Bottema, C. D. K. (2009). Limousin myostatin F94L variant affects semitendinosus tenderness. *Meat Sci.* 81, 126–131. doi: 10.1016/j.meatsci.2008.07.007

Liu, J., and Ma, D. (2014). The role of n-3 polyunsaturated fatty acids in the prevention and treatment of breast cancer. *Nutrients* 6, 5184–5223. doi: 10.3390/nu6115184

Lôbo, A. M. B. O., Guimarães, S. E. F., Paiva, S. R., Cardoso, F. F., Silva, F. F., Júnior, G. A. F., et al. (2012). Differentially transcribed genes in skeletal muscle of lambs. *Livest. Sci.* 150, 31–41. doi: 10.1016/J.LIVSCI.2012.07.027

Malau-Aduli, A. E. O., Bignell, C. W., McCulloch, R., Nichols, P. D., and Kijas, J. W., (2011). Genetic association of delta-six fatty acid desaturase single nucleotide polymorphic molecular marker and muscle long chain omega-3 fatty acids in Australian lamb. *Proceedings of the 57th International Congress of Meat Science and Technology*. pp. 141–144. Ghent, Belgium. Available at: https://isgcdata.agresearch.co.nz/ [Accessed January 25, 2019].

Nakamura, M. T., and Nara, T. Y. (2004). Structure, function, and dietary regulation of Δ6, Δ5, and Δ9 desaturases. *Annu. Rev. Nutr.* 24, 345–376. doi: 10.1146/annurev.nutr.24.121803.063211

Napolitano, F., Cifuni, G. F., Pacelli, C., Riviezzi, A. M., and Girolami, A. (2002). Effect of artificial rearing on lamb welfare and meat quality. *Meat Sci.* 60, 307–315.

Ngapo, T. M., Berge, P., Culioli, J., Dransfield, E., De Smet, S., and Claeys, E. (2002). Perimysial collagen crosslinking and meat tenderness in Belgian Blue double-muscled cattle. *Meat Sci.* 61, 91–102.

Nudda, A., McGuire, M. A., Battacone, G., and Pulina, G. (2005). Seasonal variation in conjugated linoleic acid and vaccenic acid in milk fat of sheep and its transfer to cheese and ricotta. *J. Dairy Sci.* 88, 1311–1319. doi: 10.3168/jds.S0022-0302(05)72797-1

Pfalzgraf, A., Frigg, M., and Steinhart, H. (1995). Alpha-tocopherol contents and lipid oxidation in pork muscle and adipose tissue during storage. *J. Agric. Food Chem.* 43, 1339–1342. doi: 10.1021/jf00053a039

Ponnampalam, E. N., Plozza, T., Kerr, M. G., Linden, N., Mitchell, M., Bekhit, A. E.-D. A., et al. (2017). Interaction of diet and long ageing period on lipid oxidation and colour stability of lamb meat. *Meat Sci.* 129, 43–49. doi: 10.1016/j.meatsci.2017.02.008

Ralston, J. C., Matravadia, S., Gaudio, N., Holloway, G. P., and Mutch, D. M. (2015). Polyunsaturated fatty acid regulation of adipocyte FADS1 and FADS2 expression and function. *Obesity (Silver Spring).* 23, 725–728. doi: 10.1002/oby.21035

Reardon, H. T., Hsieh, A. T., Park, W. J., Kothapalli, K. S. D., Anthony, J. C., Nathanielsz, P. W., et al. (2013). Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. *Prostaglandins. Leukot. Essent. Fatty Acids* 88, 15–19. doi: 10.1016/j.plefa.2012.02.003

Renaville, B., Prandi, A., Fan, B., Sepulcri, A., Rothschild, M. F., and Piasentier, E. (2013). Candidate gene marker associations with fatty acid profiles in heavy pigs. *Meat Sci.* 93, 495–500. doi: 10.1016/J.MEATSCI.2012.11.019

Ripoll, G., González-Calvo, L., Molino, F., Calvo, J. H., and Joy, M. (2013). Effects of finishing period length with vitamin E supplementation and alfalfa grazing

on carcass color and the evolution of meat color and the lipid oxidation of light lambs. *Meat Sci.* 93, 906–913. doi: 10.1016/j.meatsci.2012.09.017

Scollan, N. D., Dannenberger, D., Nuernberg, K., Richardson, I., MacKintosh, S., Hocquette, J.-F., et al. (2014). Enhancing the nutritional and health value of beef lipids and their relationship with meat quality. *Meat Sci.* 97, 384–394. doi: 10.1016/j.meatsci.2014.02.015

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Simopoulos, A. P. (2008). The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Exp. Biol. Med.* 233, 674–688. doi: 10.3181/0711-MR-311

Steibel, J. P., Poletto, R., Coussens, P. M., and Rosa, G. J. M. (2009). A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics* 94, 146–152. doi: 10.1016/j.ygeno.2009.04.008

Sukhija, P. S., and Palmquist, D. L. (1988). Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. *J. Agric. Food Chem.* 36, 1202–1206. doi: 10.1021/jf00084a019

Takeuchi, H., and Sugano, M. (2017). Industrial trans fatty acid and serum cholesterol: the allowable dietary level. *J. Lipids* 2017, 9751756. doi: 10.1155/2017/9751756

Terry, E. E., Zhang, X., Hoffmann, C., Hughes, L. D., Lewis, S. A., Li, J., et al. (2018). Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues. *Elife* 7. eLife. 7, e34613. doi: 10.7554/eLife.34613

Tizioto, P., Gasparin, G., Coutinho, L., Mourao, G., Mudadu, M., Souza, M., et al. (2014). MyoD1 expression levels affect meat tenderness in Nellore beef cattle. *Proceedings of 10th World Congress of Genetics Applied to Livestock Production.* pp. 601. Vancouver. Canada. Avalaible at: http://www.alice.cnptia.embrapa.br/alice/handle/doc/998063 [Accessed October 18, 2019].

Traber, M. G., and Atkinson, J. (2007). Vitamin E, antioxidant and nothing more. *Free Radic. Biol. Med.* 43, 4–15. doi: 10.1016/j.freeradbiomed.2007.03.024

Vasta, V., Pagano, R. I., Luciano, G., Scerra, M., Caparra, P., Foti, F., et al. (2012). Effect of morning vs. afternoon grazing on intramuscular fatty acid composition in lamb. *Meat Sci.* 90, 93–98. doi: 10.1016/j.meatsci.2011.06.009

Valvo, M. A., Lanza, M., Bella, M., Fasone, V., Scerra, M., Biondi, L., et al. (2005). Effect of ewe feeding system (grass v. concentrate) on intramuscular fatty acids of lambs raised exclusively on maternal milk. *Anim. Sci.* 81, 431–436. doi: 10.1079/ASC50480431

Wang, X., and Quinn, P. J. (2000). The location and function of vitamin E in membranes (review). *Mol. Membr. Biol.* 17, 143–156. doi: 10.1080/09687680010000311

Xia, J., Psychogios, N., Young, N., and Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* 37, W652–W660. doi: 10.1093/nar/gkp356

Zhang, X. Q., Jin, Y. M., Badgery, W. B., and Tana, (2017). Diet selection and n-3 polyunsaturated fatty acid deposition in lambs as affected by restricted time at pasture. *Sci. Rep.* 7, 15641. doi: 10.1038/s41598-017-15875-8

Zhu, Y., Bo, Y., and Liu, Y. (2019). Dietary total fat, fatty acids intake, and risk of cardiovascular disease: a dose-response meta-analysis of cohort studies. *Lipids Health Dis.* 18, 91. doi: 10.1186/s12944-019-1035-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dervishi, González-Calvo, Blanco, Joy, Sarto, Martin-Hernandez, Ordovás, Serrano and Calvo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 November 2019
doi: 10.3389/fgene.2019.01080

A Mini-Atlas of Gene Expression for the Domestic Goat (*Capra hircus*)

*Charity Muriuki[1,2], Stephen J. Bush[1,3], Mazdak Salavati[1,2], Mary E.B. McCulloch[1], Zofia M. Lisowski[1], Morris Agaba[4], Appolinaire Djikeng[2], David A. Hume[5] and Emily L. Clark[1,2]**

[1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [2] Centre for Tropical Livestock Genetics and Health (CTLGH), Edinburgh, United Kingdom, [3] Nuffield Department of Clinical Medicine, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom, [4] Biosciences Eastern and Central Africa - International Livestock Research Institute (BecA - ILRI) Hub, Nairobi, Kenya, [5] Mater Research Institute-University of Queensland, Woolloongabba, QLD, Australia

Goats (*Capra hircus*) are an economically important livestock species providing meat and milk across the globe. They are of particular importance in tropical agri-systems contributing to sustainable agriculture, alleviation of poverty, social cohesion, and utilisation of marginal grazing. There are excellent genetic and genomic resources available for goats, including a highly contiguous reference genome (ARS1). However, gene expression information is limited in comparison to other ruminants. To support functional annotation of the genome and comparative transcriptomics, we created a mini-atlas of gene expression for the domestic goat. RNA-Seq analysis of 17 transcriptionally rich tissues and 3 cell-types detected the majority (90%) of predicted protein-coding transcripts and assigned informative gene names to more than 1000 previously unannotated protein-coding genes in the current reference genome for goat (ARS1). Using network-based cluster analysis, we grouped genes according to their expression patterns and assigned those groups of coexpressed genes to specific cell populations or pathways. We describe clusters of genes expressed in the gastro-intestinal tract and provide the expression profiles across tissues of a subset of genes associated with functional traits. Comparative analysis of the goat atlas with the larger sheep gene expression atlas dataset revealed transcriptional similarities between macrophage associated signatures in the sheep and goats sampled in this study. The goat transcriptomic resource complements the large gene expression dataset we have generated for sheep and contributes to the available genomic resources for interpretation of the relationship between genotype and phenotype in small ruminants.

Keywords: goat, transcriptomics, RNA-Seq, gene expression, FAANG, allele-specific expression, immunity, comparative transcriptomics

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Linjie Wang, Sichuan Agricultural University, China
John C. Schwartz, Pirbright Institute, United Kingdom

***Correspondence:**
Emily L. Clark, Emily.Clark@roslin.ed.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *24 July 2019*
Accepted: *09 October 2019*
Published: *04 November 2019*

Citation:
Muriuki C, Bush SJ, Salavati M, McCulloch MEB, Lisowski ZM, Agaba M, Djikeng A, Hume DA and Clark EL (2019) A Mini-Atlas of Gene Expression for the Domestic Goat (Capra hircus). Front. Genet. 10:1080. doi: 10.3389/fgene.2019.01080

INTRODUCTION

Goats (*Capra hircus*) are an important source of meat and milk globally. They are an essential part of sustainable agriculture in low- and middle-income countries, representing a key route out of poverty particularly for women. Genomics-enabled breeding programmes for goats are currently implemented in the UK and France with breeding objectives including functional traits such as reproductive performance and disease resistance (Larroque et al., 2016; Pulina et al., 2018). The International Goat Genomics Consortium (IGGC) (http://www.goatgenome.org) has provided

extensive genetic tools and resources for goats including a 52K SNP chip (Tosser-Klopp et al., 2014), a functional SNP panel for parentage assessment and breed assignment (Talenti et al., 2018) and large-scale genotyping datasets characterising global genetic diversity (Stella et al., 2018). In 2017, a highly contiguous reference genome for goat (ARS1) was released (Bickhart et al., 2017; Worley, 2017). Advances in genome sequencing technology, particularly the development of long-read and single-molecule sequencing, meant that the ARS1 assembly was a considerable improvement in quality and contiguity from the previous whole genome shotgun assembly (CHIR_2.0) (Dong et al., 2013). In 2018, the ARS1 assembly was released on the Ensembl genome portal (Zerbino et al., 2018) (https://www.ensembl.org/Capra_hircus/Info/Index) greatly facilitating the utility of the new assembly and providing a robust set of gene models for goat.

RNA-Sequencing (RNA-Seq) has transformed the analysis of gene expression from the single-gene to the whole genome allowing visualisation of the entire transcriptome and defining how we view the transcriptional control of complex traits in livestock [reviewed in (Wickramasinghe et al., 2014)]. Using RNA-Seq, we generated a large-scale high-resolution atlas of gene expression for sheep (Clark et al., 2017). This dataset included RNA-Seq libraries from all organ systems and multiple developmental stages, providing a model transcriptome for ruminants. Analysis of the sheep gene expression atlas dataset indicated we could capture approximately 85% of the transcriptome by sampling twenty 'core' tissues and cell types (Clark et al., 2017). Given the close relationship between sheep and goats, there seemed little purpose in replicating a resource on the same scale. Our aim with the goat mini-atlas project, which we present here, was to produce a smaller, cost-effective, atlas of gene expression for the domestic goat based on transcriptionally rich tissues from all the major organ systems.

In the goat genome, there are still many predicted protein-coding and noncoding genes for which the gene model is either incorrect or incomplete, or where there is no informative functional annotation. For example, in the current goat reference genome, ARS1 (Ensembl release 97), 33% of the protein-coding genes are identified only with an Ensembl placeholder ID. Many of these unannotated genes are likely to have important functions. Using RNA-Seq data, we can annotate them and assign function (Krupp et al., 2012). With datasets of a sufficient size, genes form coexpression clusters, which can either be ubiquitous, associated with a cellular process or be cell-/tissue specific. This information can then be used to associate a function with genes coexpressed in the same cluster, a method of functional annotation known as the "guilt by association principle" (Oliver, 2000). Using this principle with the sheep gene expression atlas dataset, we were able to annotate thousands of previously unannotated transcripts in the sheep genome (Clark et al., 2017). By applying this rationale to the goat mini-atlas dataset we were able to do the same for the goat genome.

The goat mini-atlas dataset that we present here was used by Ensembl to create the initial gene build for ARS1 (Ensembl release 92). A high-quality functional annotation of existing reference genomes can help considerably in our understanding of the transcriptional control of functional traits to improve the genetic and genomic resources available, inform genomics enabled breeding programmes, and contribute to further improvements in productivity. The entire dataset is available in a number of formats to support the livestock genomics research community and represents an important contribution to the Functional Annotation of Animal Genomes (FAANG) project (Andersson et al., 2015; FAANG, 2017; Harrison et al., 2018).

This study is the first global analysis of gene expression in goats. Using the goat mini-atlas dataset, we describe large clusters of genes associated with the gastrointestinal tract and macrophages. Species specific differences in response to disease, or other traits, are likely to be reflected in gene expression profiles. Sheep and goats are both small ruminant mammals and are similar in their physiology. They also share susceptibility to a wide range of viral, bacterial, parasitic, and prion pathogens, including multiple potential zoonoses (Sherman, 2011), but there have been few comparisons of relative susceptibility or pathology between the species to the same pathogen nor the nature of innate immunity. To reveal transcriptional similarities and differences between sheep and goats, we have performed a comparative analysis of gene expression by comparing the goat mini-atlas dataset with a comparable subset of data from the sheep gene expression atlas (Clark et al., 2017). We also use the goat mini-atlas dataset to examine the expression of candidate genes associated with functional traits in goats and link these with allele-specific expression (ASE) profiles across tissues, using a robust methodology for ASE profiling (Salavati et al., 2019). The goat mini-atlas dataset and the analysis we present here provide a foundation for identifying the regulatory and expressed elements of the genome that are driving functional traits in goats.

METHODS

Animals

Tissue and cell samples were collected from six male and one female neonatal crossbred dairy goats at six days old. The experimental design was based on sample availability at the time of the study. The goats were sourced from one farm and samples were collected at a local abattoir within 1 h of euthanasia.

Tissue Collection

The tissue samples were excised postmortem within 1 h of death, cut into 0.5cm diameter segments, and transferred into RNAlater (Thermo Fisher Scientific, Waltham, USA) and stored at 4°C for short-term storage. Within one week, the tissue samples were removed from the RNAlater, transferred to 1.5ml screw cap cryovials, and stored at -80°C until RNA isolation. Alveolar macrophages (AMs) were isolated from two male goats by broncho-alveolar lavage of the excised lungs using the method described for sheep in (Clark et al., 2017), except using 20% heat-inactivated goat serum (G6767, Sigma Aldrich), and stored in TRIzol (15596018; Thermo Fisher

Scientific) at -80°C for RNA extraction. Similarly, bone marrow cells (BMCs) were isolated from 10 ribs from 3 male goats and frozen down for subsequent differentiation and stimulation with lipopolysaccharide (LPS) using the method described in (Clark et al., 2017; Young et al., 2018). Bone marrow derived macrophages (BMDMs) were obtained by culturing BMCs for 10 days in complete medium: RPMI 1640, Glutamax supplement (35050–61; Invitrogen), 20% heat inactivated goat serum (G6767; Sigma Aldrich), penicillin/streptomycin (15140, Invitrogen), and in the presence of recombinant human CSF-1 (rhCSF-1: 10^4 U/ml; a gift of Chiron, Emeryville, CA) on T75 polystyrene tissue culture treated plates (156499; Thermo Fisher Scientific) at a density of $2.0x10^6$cells/ml. On day 11, BMDMs were transferred to 6-well cell culture treated multidishes (140675; Thermo Fisher Scientific). The following day, they were stimulated with LPS from *Salmonella enterica* serotype minnesota Re 595 (L9764; Sigma-Aldrich) at a final concentration of 100 ng/ml, then transferred into TRIzol (15596018; Thermo Fisher Scientific) at 0, 7h post LPS treatment, and stored at -80°C for RNA extraction.

Details of all the samples collected are included in **Table 1**.

RNA Extraction

RNA was extracted from tissues and cells as described in (Clark et al., 2017). For each RNA extraction from tissues, approximately 60mg of tissue was processed. Tissue samples were homogenised on a Precellys Tissue Homogeniser (Bertin Instruments; Montigny-le-Bretonneux, France) at 5000 rpm for 20 s with CK14 (432–3751; VWR, Radnor, USA) tissue homogenising ceramic beads in 1ml of TRIzol (15596018; Thermo Fisher Scientific). Cell samples were collected atthe point of isolation into TRIzol (15596018; Thermo Fisher Scientific), stored at -80°C, thawed, and then mixed by pipetting to homogenise. To allow sufficient time for complete dissociation of the nucleoprotein complex, homogenised (cell/tissue) samples were incubated at room temperature for 5 min. After 5 min, 200µl BCP (1-bromo-3-chloropropane) (B9673; Sigma Aldrich) was added and the sample was shaken vigorously for 15 s and incubated at room temperature for a further 3 min. The homogenised sample was then centrifuged for 15 min at 12,000 x *g*, at 4°C for 3 min, to separate the upper clear aqueous layer. This clear upper layer was then column purified to remove DNA and trace phenol using a RNeasy Mini Kit (74106; Qiagen Hilden, Germany) following the manufacturer's instructions (RNeasy Mini Kit Protocol: Purification of Total RNA from Animal Tissues, from step 5 onwards). An on-column DNase treatment was performed using the Qiagen RNase-Free DNase Set (79254; Qiagen Hilden, Germany). The sample was eluted in 30ul of RNase free water and stored at -80°C prior to QC and library preparation. To ensure RNA integrity (RIN^e) was of $RIN^e > 7$ samples were run on an Agilent 2200 TapeStation System (Agilent Genomics, Santa Clara, USA). RIN^e and other quality control metrics for the RNA samples are included in **Supplementary Table S1**.

TABLE 1 | Details of samples included in the goat mini-atlas.

Tissue/cell type	Organ system	No. of replicates	Sex
Adrenal gland	Endocrine	4	male
Alveolar macrophage	Immune	2	male
BMDM - LPS (0 h)	Immune	3	male
BMDM + LPS (7 h)	Immune	3	male
Cerebellum	Nervous system	2	male
Colon large	GI tract	4	male
Fallopian tube	Reproductive system	1	female
Frontal lobe cortex	Nervous system	2	male
Ileum and Peyer's patches	GI tract	2	male
Kidney cortex	Endocrine	4	male
Liver	Endocrine	4	male
Ovary	Reproductive system	1	female
Rumen	Gastrointestinal tract	2	male
Skeletal muscle - longissimus dorsi	Musculo-skeletal	3	male
Skin	Integumentary	4	male
Spleen	Immune	3	male
Testes	Reproductive system	4	male
Thymus	Immune	4	male
Uterine horn	Reproductive system	1	female
Uterus	Reproductive system	1	female

RNA-Sequencing

RNA-Seq libraries were sequenced on the Illumina HiSeq 4000 sequencing platform (Illumina, San Diego, USA) and generated by Edinburgh Genomics (Edinburgh Genomics, Edinburgh, UK). Strand-specific paired-end reads with a fragment length of 75bp were generated for each sample using the standard Illumina TruSeq mRNA library preparation protocol (poly-A selected) (Ilumina; Part: 15031047 Revision E). Libraries were sequenced at a depth of either >30 million reads per sample for the tissues and AMs, or >50 million reads per sample for the BMDMs.

Data Processing

The RNA-Seq data processing methodology and pipelines are described in detail in (Clark et al., 2017). Briefly, for each tissue, a set of expression estimates, as transcripts per million (TPM), were obtained. These estimates were obtained using the alignment-free (technically, "pseudo-aligning") transcript quantification tool Kallisto (Bray et al., 2016), the accuracy of which is dependent on a high quality reference transcriptome (index). We used a "two-pass" approach to generate this index in order to ensure we used an accurate set of gene expression estimates.

To generate the index, we initially ran Kallisto on all samples using as its index the ARS1 reference transcriptome available from Ensembl (ftp://ftp.ensembl.org/pub/release-95/fasta/capra_hircus/cdna/Capra_hircus.ARS1.cdna.all.fa.gz). The resulting data was then parsed to revise this index. This was for two reasons: i) so that we included in the second index, those transcripts that were missing but should have been present (i.e. due to incompleteness in the reference annotation), and ii) to remove transcripts that were present but should not have been (i.e., where a spurious model was present in the reference annotation). For i), we

obtained the subset of reads that Kallisto could not (pseudo) align, assembled those *de novo* into putative transcripts, then retained each transcript only if it could be robustly annotated. We considered annotation robust when a transcript encoded a protein similar to one of known function and had coding potential. For ii), we identified those transcripts in the reference transcriptome for which no evidence of expression could be found in any of the samples from the goat mini-atlas and discarded them. This revised index was used for a second "pass" with Kallisto to generate expression level estimates with higher-confidence.

We complemented the Kallisto alignment-free method with a reference-guided alignment-based approach to RNA-Seq processing, using the HISAT aligner (Kim et al., 2015) and StringTie assembler (Pertea et al., 2015). This approach was highly accurate when mapping to the (ARS1) annotation on NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1/GCF_001704415.1_ARS1_rna.fna.gz), precisely reconstructing almost all exon (96%) and transcript (76%) models (**Supplementary Table S2**). We used the HISAT/StringTie output to validate the set of transcripts used to generate the Kallisto index. HISAT/StringTie, unlike Kallisto and other alignment-free methods, can be used to identify novel transcript models, particularly for ncRNAs, which we have described separately in (Bush et al., 2018b). Details of all novel transcript models detected are included in **Supplementary Table S3.**

Data Validation

To identify any spurious samples which could have been generated during sample collection, RNA extraction, or library preparation, we generated a sample-to-sample correlation of the gene expression estimates from Kallisto, in Graphia Professional (Kajeka Ltd, Edinburgh, UK).

Network Cluster Analysis

Network cluster analysis of the goat gene mini-atlas dataset was performed using Graphia Professional (Kajeka Ltd, Edinburgh, UK) (Livigni et al., 2018). Briefly, by calculating a Pearson correlation matrix for both gene-to-gene and sample-to-sample comparisons, and filtering to remove relationships where $r <$ 0.83, we were able to determine similarities between individual gene expression profiles. A network graph was constructed by connecting the nodes (transcripts) with edges (where the correlation exceeded the threshold value). Network graphs were interpreted by applying a Markov Cluster algorithm (MCL) at an inflation value/cluster granularity of 2.2 (Freeman et al., 2007). The granularity of the network graph was manually curated in order to reach a biologically relevant number of interaction nodes and cluster numbers. This approach was iteratively applied to several correlation coefficient thresholds for comparison prior to clustering, as previously described in Freeman et al., 2007, Clark et al., 2017. A suitable correlation threshold of 0.83 was chosen and the local structure of the graph was then examined visually. Transcripts with related functions clustered together forming sets of tightly interlinked nodes. The principle of "guilt by association" was then applied, to infer the function of unannotated genes from genes within the same cluster (Oliver, 2000). Clusters 1 to 30 were assigned a functional "class" based on whether transcripts within a cluster shared a similar biological function according to GO term enrichment using the Bioconductor package "topGO" (Alexa and Rahnenfuhrer, 2010).

Comparative Analysis of Gene Expression in Macrophages in Sheep and Goats

To compare transcriptional differences in the immune response between the two species, we focused our analysis on the macrophage populations (AMs and BMDMs). For this analysis, we used a subset of data from our sheep gene expression atlas for AMs and BMDMs (+/- LPS) from three male sheep (Clark et al., 2017) (**Supplementary Dataset S1**).

For AMs, we compared the gene level expression estimates from the two male goats and the three male sheep using edgeR v3.20.9 (Robinson et al., 2010). Only genes with the same gene name in both species, expressed at a raw read count of more than 10, FDR < 10%, an FDR adjusted p-value of <0.05, and Log2FC of > = 2, in both goat and sheep, were included in the analysis.

Differential expression analysis using edgeR (Robinson et al., 2010) was also performed for sheep and goat BMDMs (+/-) LPS separately, using the filtration criteria described above for AMs, to compile a list of genes for each species that were up or down regulated in response to LPS. These lists were then compared using the R package dplyr (Wickham et al., 2018) with system query language syntax. Each list was merged based on GENE_ID using the *inner_join* function to only return the observations that overlapped between goat and sheep (i.e., genes which had corresponding annotations in both species).

A dissimilarity index (Dis_Index) was then calculated by taking the absolute difference (ABS) of the Log2 fold change (Log2FC) between sheep and goat using the formula:

$$\mathrm{ABS}(\mathrm{Log2FC}^{\mathrm{Sheep}} - \mathrm{Log2FC}^{\mathrm{Goat}})$$

A high Dis_Index indicated that a gene was differently regulated in goat and sheep.

Allele-Specific Expression

To measure allele-specific expression (ASE), across tissues and cell-types from the goat mini-atlas, we used the method described in (Salavati et al., 2019). Briefly, BAM files from the RNA-Seq data were mapped to the ARS1 top level DNA fasta track from Ensembl v96, using HISAT2 as described in (Clark et al., 2017). Any reference mapping bias was removed using WASP v0.3.1 (van de Geijn et al., 2015) and the resultant BAM files processed using the Genome Analysis Tool Kit (GATK) to produce individual VCF files. The ASEreadCounter tool in GATK v3.8 was used to obtain raw counts of the allelic expression profile in the dataset. These raw counts were then tested for imbalance (using a modified negative-beta bionomial test at gene level) at all heterozygote loci (i.e., ASE = Counts $_{\mathrm{RefAllele}}$/(Counts $_{\mathrm{RefAllele}}$+ Counts $_{\mathrm{AltAllele}}$) within the boundaries of the gene using the R package GeneiASE (Edsgärd et al., 2016).

RESULTS AND DISCUSSION

Scope of the Goat Mini-Atlas Dataset, Sequencing Depth, and Coverage

The goat mini-atlas dataset includes 54 mRNA-Seq (poly-A selected) 75bp paired-end libraries. Details of the libraries generated including the age and sex of the animals, the tissues and cell types sampled, and the number of biological replicates per sample are summarised in **Table 1**. Gene level expression estimates, for the goat mini-atlas, are provided as unaveraged (**Supplementary Dataset S2**) and averaged across biological replicates (**Supplementary Dataset S3**) files.

Approximately, $8.7x10^8$ paired end sequence reads were generated in total. Following data processing with Kallisto (Bray et al., 2016), a total of 18,528 unique protein coding genes had detectable expression (TPM > 1), representing 90% of the reference transcriptome (Bickhart et al., 2017). From the set of 17 tissues and 3 cell types we sampled, we were able to detect approximately 90% of protein coding genes providing proof of concept that the mini-atlas approach is useful for global analysis of transcription. The average percentage of transcripts detected per tissue or cell type was 66%, ranging from 54% in alveolar macrophages, which had the lowest to 72% in testes, which had the highest. The percentage of protein coding genes detected per tissue is included in **Table 2**. Although we included uterine horn as well as uterus and both stimulated and unstimulated BMDMs, our analysis suggests that including only one tissue/cell of a similar type would be the most economical approach to generating a mini-atlas of gene expression for functional annotation.

TABLE 2 | The percentage of protein coding genes detected per tissue in the goat mini-atlas dataset.

Tissue	Average no. of protein-coding genes expressed (TPM > 1) in this tissue	% of protein-coding genes expressed (TPM > 1) in this tissue
Adrenal gland	14585	68.34
Alveolar macrophage	11533	54.04
BMDM - LPS (0 h)	13253	62.1
BMDM + LPS (7 h)	13042	61.11
Cerebellum	14959	70.09
Colon large	14736	69.04
Fallopian tube	14390	67.42
Frontal lobe cortex	14757	69.14
Ileum and Peyer's patches	15268	71.54
Kidney cortex	15223	71.33
Liver	13497	63.24
Ovary	14251	66.77
Rumen	13642	63.92
Skeletal muscle - longissimus dorsi	12276	57.52
Skin	14892	69.77
Spleen	14659	68.68
Testes	15359	71.96
Thymus	14484	67.86
Uterine horn	14298	66.99
Uterus	14298	66.99

Approximately, 2,815 (13%) of the total 21,343 protein coding genes in the goat reference transcriptome had no detectable expression in the goat mini-atlas dataset. These transcripts are likely to be either tissue specific to tissues and cell-types that were not sampled here (including lung, heart, pancreas, and various endocrine organs), rare, or not detected at the depth of coverage used. The large majority of these transcripts were detected in the much larger sheep atlas, and their likely expression profile can be inferred from the sheep. In addition, for the goat mini-atlas unlike the sheep gene expression atlas, we only included neonatal animals so transcripts that were highly developmental stage-specific in their expression pattern would also not be detected. A list of all undetected genes is included in **Supplementary Table S4** and undetected transcripts in **Supplementary Table S5**.

Gene Annotation

The proportion of transcripts per biotype (lncRNA, protein coding, pseudogene, etc), with detectable expression (TPM >1) in the goat mini-atlas relative to the ARS1 reference transcriptome, on Ensembl is summarised at the gene level in **Supplementary Table S6** and at the transcript level in **Supplementary Table S7**. Of the 21,343 protein coding genes in the ARS1 reference transcriptome, 7036 (33%) had no informative gene name. Whilst the Ensembl annotation will often identify homologues of a goat gene model, the automated annotation genebuild pipeline used to assign gene names and symbols is conservative. Using the annotation pipeline we described in (Clark et al., 2017), we were able to use the goat mini-atlas dataset to assign an informative gene name to 1114 previously unannotated protein coding genes in ARS1. These genes were annotated by reference to the NCBI nonredundant (nr) peptide database v94 (Pruitt et al., 2007). A shortlist containing a conservative set of gene annotations to HGNC (HUGO Gene Nomenclature Committee) gene symbols is included in **Supplementary Table S8**. **Supplementary Table S9** contains the full list of genes annotated using the goat mini-atlas dataset and our annotation pipeline. Many unannotated genes can be associated with a gene description, but not necessarily an HGNC symbol; these are also listed in **Supplementary Table S10**.

Network Cluster Analysis

Network cluster analysis of the goat gene expression atlas was performed using Graphia Professional (Kajeka Ltd, Edinburgh UK), a network visualisation tool (Livigni et al., 2018). The goat mini-atlas unaveraged TPM estimates (**Supplementary Dataset S2**) were used for network cluster analysis. We first generated a sample-to-sample graph (r = 0.75, MCL = 2.2) **Supplementary Figure S1**, which verified that the correlation between biological replicates was high and that none of the samples were spurious. We then generated a gene-to-gene network graph (**Figure 1**), with a Pearson correlation coefficient of r = 0.83, that comprised 16,172 nodes (genes) connected by 1,574,259 edges. The choice of Pearson correlation threshold is optimised within the Graphia program to maximise the number of nodes (genes) included whilst minimising the number of edges (Freeman et al., 2007). By applying the MCL (Markov Clustering) algorithm at an inflation value (which determines cluster

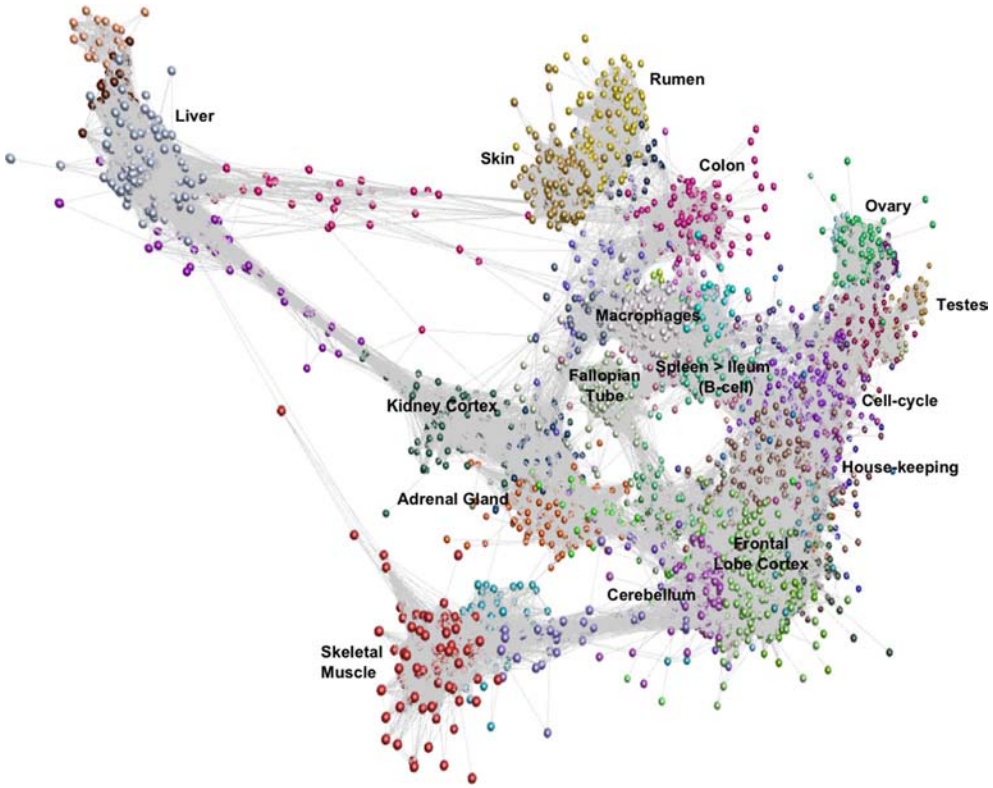


FIGURE 1 | Gene-to-gene network graph of the goat mini-atlas dataset. Each node represents a gene and each edge represents correlations between individual measurements above the set threshold. The graph comprised 16,172 nodes (genes) and 1,574,259 edges (Pearson correlations ≥ 0.83), Markov Cluster algorithm (MCL) inflation = 2.2, and Pearson Product Correlation Co-efficient = 0.83. (> indicates decreasing expression profile).

granularity) of 2.2, the gene network graph separated into 75 distinct coexpression clusters, with the largest cluster (cluster 1) comprising of 1795 genes. Genes found in the top 30 largest clusters are listed in **Supplementary Table S11**. Clusters 1 to 20 (numbered in order of size, largest to smallest) were annotated manually and assigned a functional "class" (**Table 3**). These functional classes were assigned based on GO term enrichment (Alexa and Rahnenfuhrer, 2010) for molecular function and biological process (**Supplementary Table S12**). Assignment of functional class was further validated by visual inspection of expression pattern and comparison with functional groupings of genes observed in the sheep gene expression atlas (Clark et al., 2017).

The largest of the clusters (Cluster 1) contained 1,795 genes that were almost exclusively expressed in the central nervous system (cortex, cerebellum) reflecting the high transcriptional activity and complexity in the brain. Significant GO terms for cluster 1 included cognition ($p = 4.6x10^{-17}$) and synaptic transmission ($p = 2.5x10^{-30}$). Other tissue-specific clusters, e.g., 4 (liver), 6 (testes), 7 (skin/rumen), 14 (adrenal), and 17 (kidney), were similarly enriched for genes associated with known tissue-specific functions. In each case, the likely function of unannotated protein-coding genes within these clusters could be inferred by association with genes of known function that share the same cell or tissue specific expression pattern. Cluster 9 showed a high level of tissue specificity and included genes associated with skeletal muscle function and development including *MSTN* which encodes a protein that negatively regulates skeletal muscle cell proliferation and differentiation (Wang et al., 2012). Several myosin light and heavy chain genes (e.g., *MYH1* and *MYL1*) and transcription factors that are specific to muscle including (*MYOG* and *MYOD1*) were also found in cluster 9. GO terms for muscle were enriched in cluster 9, e.g., muscle fiber development ($p = 3.8x10^{-13}$) and structural constituent of muscle ($p = 1.8x10^{-11}$). Genes expressed in muscle are of particular biological and commercial interest for livestock production and represent potential targets for gene editing (Yu et al., 2016). Cluster 8 was also highly tissue specific and included genes expressed in the fallopian tube with enriched GO terms for cilium movement ($p = 1.4x10^{-15}$) and cilium organization ($p = 2.3x10^{-15}$). A motile cilia cluster was identified in the fallopian tube in the sheep gene expression atlas (Clark et al., 2017) and a similar cluster was enriched in chicken in the trachea (Bush et al., 2018a). The goat mini-atlas also included several clusters that were enriched for immune tissues and cell types and we have based our analysis in part upon the premise that the greatest differences between small ruminant species likely involve the immune system.

Gene Expression in the Neonatal Gastrointestinal Tract

Three regions of the gastrointestinal (GI) tract were sampled; the ileum, colon, and rumen. These regions formed distinct

TABLE 3 | Annotation of the 20 largest network clusters in the goat mini-atlas dataset (> indicates decreasing expression profile).

Cluster ID	Number of genes	Profile description	Class	Enriched GO terms
1	1795	Cortex > cerebellum	Brain	cognition, neurotransmitter transport, synaptic transmission
2	1395	Thymus > Spleen > Ileum	Cell-Cycle	DNA-dependent DNA replication, DNA repair
3	795	General	House Keeping	mRNA processing, regulation of RNA splicing
4	505	Liver	Oxidative-Phosphorylation	oxidation-reduction process, fatty acid oxidation
5	494	General	House Keeping	RNA binding, nucleolus
6	481	Testes	Male Reproduction	male meiosis, spermatogenesis
7	449	Skin > Rumen	Epithelial	skin morphogenesis, keratinocyte differentiation
8	374	Fallopian Tube	Motile Cilia	motile cilium, ciliary basal body
9	351	Skeletal muscle	Muscle	muscle fibre development, motor activity
10	337	Spleen > Ileum	Immune	immune response, B-cell activation, cytokine activity
11	290	Macrophages	Immune	response to lipopolysaccharide, phagocytic vesicle
12	241	Colon Large	Gastrointestinal tract	microvillus, actin filament bundle
13	226	Rumen > Skin	Gastrointestinal/Epithelial	epidermis development, chloride channel activity
14	219	Adrenal Gland	Endocrine	oxidation-reduction process, sterol metabolic process
15	211	BMDMs	Fibroblasts	collagen binding, positive regulation of fibroblast proliferation
16	134	General	Ribosomal	ribosomal large subunit biogenesis, ribosome
17	133	Kidney Cortex	Mesoendonephric organogenesis	sodium ion homeostasis, skeletal system morphogenesis
18	119	Ovary	Oogenesis	growth factor activity, nucleosome disassembly
19	113	Ileum > Spleen > Thymus	Immune	B-cell proliferation, cytokine activity
20	108	Uterus, Uterine Horn	Organogenesis	tissue remodelling, bone morphogenesis

clusters in the network graph. The genes comprising these clusters were highly correlated with the physiology of the tissues. Goats are ruminant mammals and, at one-week of age (when tissues were collected), the rumen is vestigial. Even at this early stage of development, the typical epithelial signature of the rumen (Xiang et al., 2016a; Xiang et al., 2016b) was observed. Genes coexpressed in the rumen (clusters 7 and 13 – **Table 3**) were typical of a developing rumen epithelial signature (Bush et al., 2019) and were associated with GO terms for epidermis development ($p = 0.00016$), keratinocyte differentiation ($p = 1.5x10^{-14}$), and skin morphogenesis ($p = 8.2x10^{-6}$). Large colon (cluster 12) included several genes associated with GO terms for microvillus organization ($p = 1x10^{6}$) and microvillus ($p = 6.3x10^{6}$) including *MYO7B* which is found in the brush border cells of epithelial microvilli in the large intestine. The microvilli function as the primary surface of nutrient absorption in the gastrointestinal tract, and as such numerous phospholipid-transporting ATPases and solute carrier genes were found in the large colon cluster.

Throughout the GI tract, there was a strong immune signature, similar to that observed in neonatal and adult sheep (Bush et al., 2019), which was greatest in clusters 10 and 19 (**Table 3**) where expression was high in the ileum and Peyer's patches, thymus, and spleen. Cluster 10 had a more general immune related profile with higher expression in the spleen and significant GO terms associated with cytokine receptor activity ($p = 1.3x10^{-8}$) and T-cell receptor complex ($p = 0.00895$). Several genes involved in the immune and inflammatory response were found in cluster 10 including *CD74*, *IL10*, and *TLR10*. The expression pattern for cluster 19 was associated with B-cells including GO terms for B-cell proliferation ($p = 1.4x10^{-7}$), positive regulation of B-cell activation ($p = 4.9x10^{-6}$), and cytokine activity ($p = 0.0051$). Genes associated with the B-cell receptor complex *CD22*, *CD79B*, *CD180*, and *CR2*, and interleukins *IL21R* and *IL26* were expressed in cluster 19 (Treanor, 2012). This reflects the fact that we sampled the Peyer's patch with the ileum, which is a primary lymphoid organ of B-cell development in ruminants (Masahiro et al., 2006).

Each of the GI tract clusters included genes associated with more than one cell type/cellular process. This complexity is a consequence of gene expression patterns from the lamina propria, one of the three layers of the mucosa. The lamina propria lies beneath the epithelium along the majority of the GI tract and comprises numerous different cell types from endothelial, immune and connective tissues (Ikemizu et al., 1994). This gene expression pattern, which is also observed in sheep (Clark et al., 2017; Bush et al., 2019) and pigs (Freeman et al., 2012), highlights the complex multidimensional physiology of the ruminant GI tract.

Macrophage-Associated Signatures

A strong immune response is vitally important to neonatal mammals. Macrophages constitute a major component of the innate immune system acting as the first line of defense against invading pathogens and coordinating the immune response by triggering antimicrobial responses and other mediators of the inflammatory response (Hume, 2015). Several clusters in the goat mini-atlas exhibited a macrophage-associated signature. Cluster 11 (**Table 3**) contained several macrophage marker genes, including *CD68* which is expressed in AMs and BMDMs. The cluster includes the macrophage growth factor, *CSF1*, indicating that as in sheep (Clark et al., 2017), pigs (Freeman et al., 2012), and humans (Schroder et al., 2012) but in contrast to mice, according to the results of this study goat macrophages are autocrine for their own growth factor. GO terms associated with cluster 11 included phagocytosis ($p = 3.5x10^{-10}$), inflammatory response ($p = 1.4x10^{-8}$), and cytokine

receptor activity (p = 0.00031). Many of the genes that were up-regulated in AMs in cluster 11, including C-type lectins *CLEC4A* and *CLEC5A*, have been shown to be down regulated in sheep (Clark et al., 2017; Bush et al., 2019), pigs (Freeman et al., 2012), and humans (Baillie et al., 2017) in the wall of the intestine. This highlights functional transcriptional differences in macrophage populations. AMs respond to microbial challenge as the first line of defense against inhaled pathogens. In contrast, macrophages in the intestinal mucosa down-regulate their response to microorganisms as a continuous inflammatory response to commensal microbes would be undesirable.

Cluster 11 (**Table 3**) also included numerous proinflammatory cytokines and chemokines which were up-regulated following challenge with lipopolysaccharide (LPS). Response to LPS was also reflected in several significant GO terms associated with this cluster including, cellular response to lipopolysaccharide ($p = 5.8x10^{-10}$), and cellular response to cytokine stimulus ($p = 9.5x10^{-8}$). C-type lectin *CLEC4E*, which is known to be involved in the inflammatory response (Baillie et al., 2017), interleukin genes such as *IL1B* and *IL27*, and *ADGRE1* were all highly inducible by LPS in BMDMs. *ADGRE1* (*EMR1,F4/80*) is a monocyte-macrophage marker involved in pattern recognition which exhibits interspecies variation both in expression level and response to LPS stimulation (Waddell et al., 2018). Based upon RNA-Seq data, ruminant genomes were found to encode a much larger form of *ADGRE1* than monogastric species, with complete duplication of the extracellular domain [44].

Comparative Analysis of Macrophage-Associated Transcriptional Responses in Sheep and Goats

Transcriptional differences are linked to species-specific variation in response to disease, and have been widely documented in livestock (Bishop and Woolliams, 2014). For instance, ruminants differ in their response to a wide range of economically important pathogens. Variation in the expression of *NRAMP1* (*SLC11A1*) is involved in the response of sheep and goat to Johne's disease (Cecchi et al., 2017). Similarly, resistance to *Haemonchus contortus* infections in sheep and goats is associated with a stronger Th2-type transcriptional immune response (Gill et al., 2000; Alba-Hurtado and Munoz-Guzman, 2013). To determine whether goats and sheep differ significantly in immune transcriptional signatures, we performed a comparative analysis of the macrophage samples from the goat mini-atlas and those included in our gene expression atlas for sheep (Clark et al., 2017). One caveat to this analysis that should be noted is that the sheep and goat samples were unfortunately not age-matched and as such differences in gene expression could be an effect of developmental stage rather than species-specific differences. However, as macrophage samples from both species were kept in culture prior to collection and analysis, we would expect the effect of developmental stage to be minimal.

We performed differential analysis of genes expressed in goat and sheep AMs (**Supplementary Table S13**). The top 25 genes up- and down-regulated in goat relative to sheep based on log2FC are shown in **Figure 2**. Several genes involved in

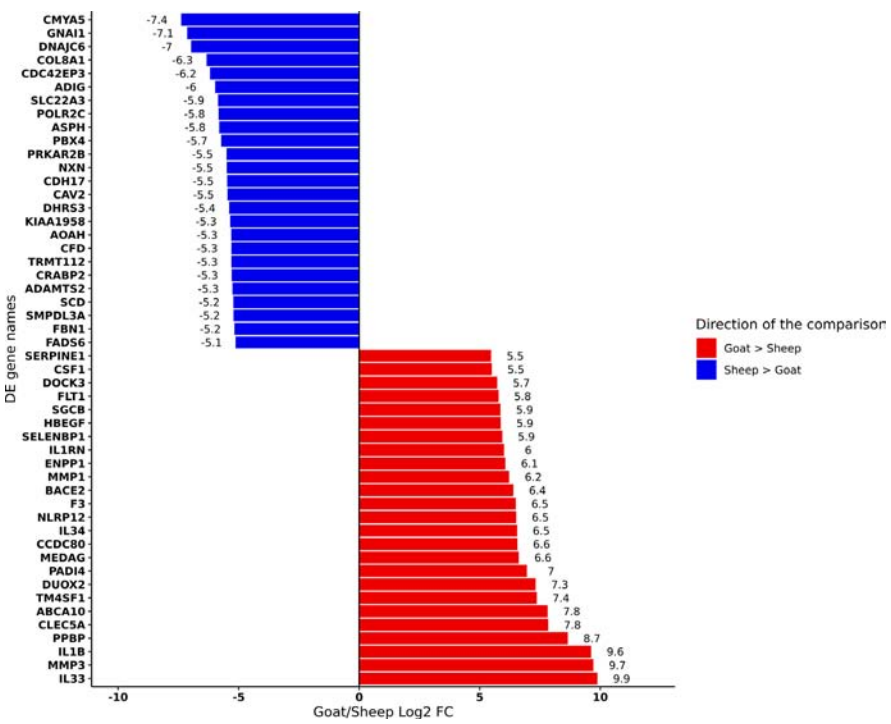


FIGURE 2 | Differentially expressed genes (FDR < 10%) between goat and sheep alveolar macrophages. The top 25 up-regulated in goat relative to sheep (red) and the top 25 down-regulated in goat relative to sheep (blue) are shown.

the inflammatory and immune response including interleukins *IL33* and *IL1B* and C-type lectin *CLEC5A* were up-regulated in goat AMs relative to sheep. In contrast, those that were down regulated in goat relative to sheep did not have an immune function but were associated with more general physiological processes. This may reflect species-specific differences but could also indicate that the immune response in AMs is age-dependent, i.e., neonatal animals exhibit a primed immune response while a more subdued response is exhibited by adult sheep whose adaptive immunity has reached full development.

Using differential expression analysis (Robinson et al., 2010), we also compared the gene expression estimates for sheep and goat BMDMs (+/-) LPS to compile a list of genes for each species that were up or down regulated in response to LPS (**Supplementary Table S14A** goat and **Supplementary Table S14B** sheep). These lists were then merged using the methodology described above (see Methods section) to highlight genes that differed in their response to LPS between the two species. In total, 188 genes exhibited significant differences between goats and sheep (FDR < 10%, Log2FC> = 2) in response to LPS (**Supplementary Table S15**). The genes which showed the highest level of dissimilarity in response to LPS between goats and sheep (Dis_Index> = 2) are illustrated in **Figure 3**. Several immune genes were upregulated in both goat and sheep BMDMs in response to LPS stimulation but differed in their level of induction between the two species (top right quadrant **Figure 3**). *IL33*, *IL36B*, *PTX3*, *CCL20, CSF3*, and *CSF2* for example, exhibited higher levels of induction in sheep BMDMs relative to goat, and vice versa for *ICAM1*, *IL23A*, *IFIT2*, *TNFSF10*, and *TNFRSF9*. Several genes were upregulated in sheep but downregulated in goat BMDMs (e.g., *KIT*) (top left quadrant **Figure 3**), and upregulated in goat, but downregulated in sheep (e.g., *IGFBP4*) (bottom right quadrant **Figure 3**).

Overall, the transcriptional patterns in BMDMs stimulated with LPS were broadly similar between the two species. Although, further experiments using qPCR to measure the expression of candidate genes in age-matched animals would be required to validate the observed expression patterns. With this caveat in mind, some interesting differences in individual genes were observed that could contribute to species-specific responses to infection. For instance, *IL33* and *IL23A* both exhibited a higher level of induction in sheep BMDMs after stimulation with LPS relative to goat (**Figure 3**). In humans, *IL33* has a protective role in inflammatory bowel disease by inducing a Th2 immune response (Lopetuso et al., 2013). An enhanced Th2 response, which accelerates parasite expulsion, has been associated with *H. contortus* resistance in sheep (Alba-Hurtado and Munoz-Guzman, 2013). Conversely, higher expression of *IL23A* is associated with susceptibility to *Teladorsagia circumcincta* infection (Gossner et al., 2012). Little is known about the function of *IL33* and *IL23A* in goats. They are members of the interleukin-1 family which play a central role in the regulation of immune and inflammatory response to infection (Dinarello, 2018). Given the similarities in their expression patterns, it is reasonable to assume that these genes are regulated in a similar manner to sheep and involved in similar biological pathways. As such, they would be suitable candidate genes to investigate

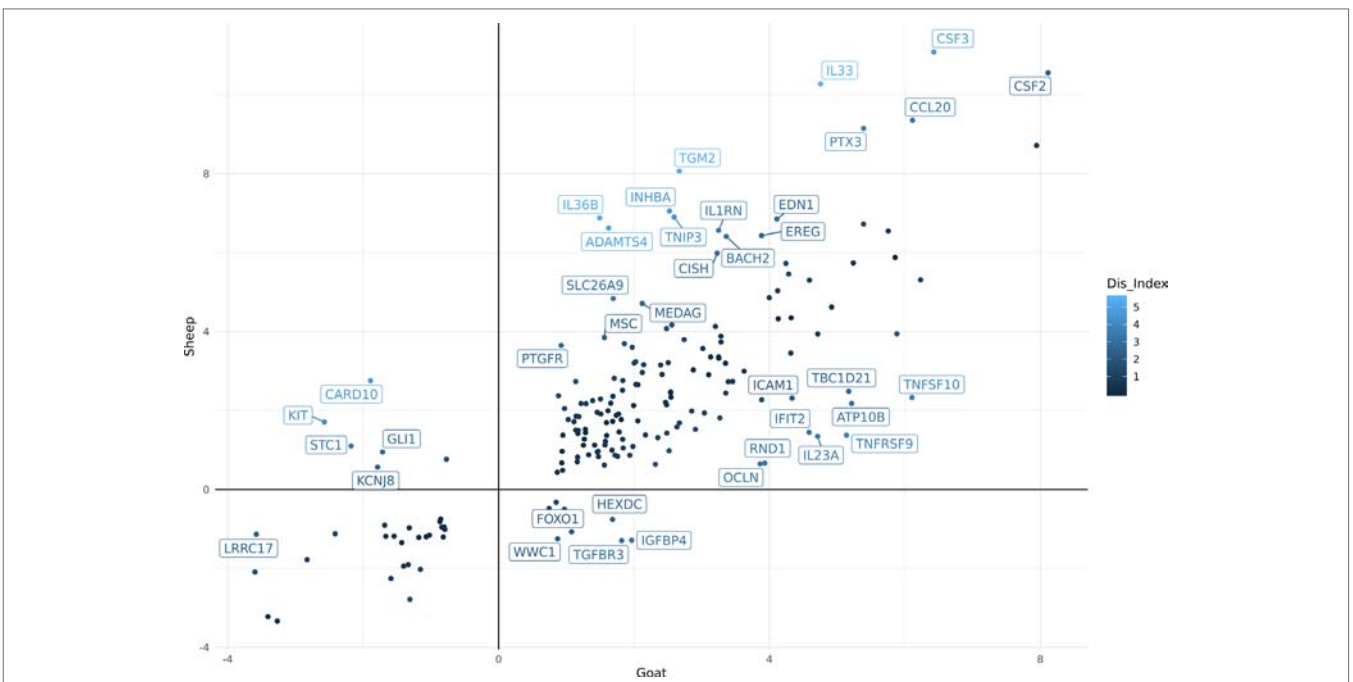


FIGURE 3 | Comparative analysis of differentially expressed genes (FDR < 10%, Log2FC> = 2) in goat and sheep bone marrow derived macrophage (BMDM). The genes which showed the highest level of dissimilarity in response to lipopolysaccharide (LPS) between goats and sheep (Dis_Index> = 2) are shown. Top right quadrant: genes that were up-regulated in both goat and sheep but differed in their level of induction between the two species. Top left quadrant: genes that were up-regulated in sheep but down-regulated in goat. Bottom right quadrant: genes up-regulated in goat, but down-regulated in sheep.

further to determine if they underlie species-specific variation in susceptibility to pathogens (Bishop and Stear, 2003; Bishop and Morris, 2007).

Expression Patterns of Genes Associated With Functional Traits in Goats

The goat mini-atlas dataset is a valuable resource that can be used by the livestock genomics community to examine the expression patterns of genes of interest that are relevant to ruminant physiology, immunity, welfare, production, and adaptation/resilience particularly in tropical agri-systems. The mini-atlas provides a resource of tissue-specific expression profiles for each gene that could be used to help determine which tissues to prioritise, for an expression QTL study, for example. Several genes, associated with functional traits in goats, have been identified using genome wide association studies (GWAS). Insulin-like growth factor 2 (*IGF2*), for example, is associated with growth rate in goats (Burren et al., 2016), and was highly expressed in tissues with a metabolic function including, kidney cortex, liver, and adrenal gland (**Figure 4A**). As expected, expression of myostatin (*MSTN*), which encodes a negative regulator of skeletal muscle mass, was highest in skeletal muscle in comparison with the other tissues (**Figure 4B**). *MSTN* is a target for gene-editing in goats to promote muscle growth (e.g., Yu et al., 2016). Expression of genes associated with fecundity

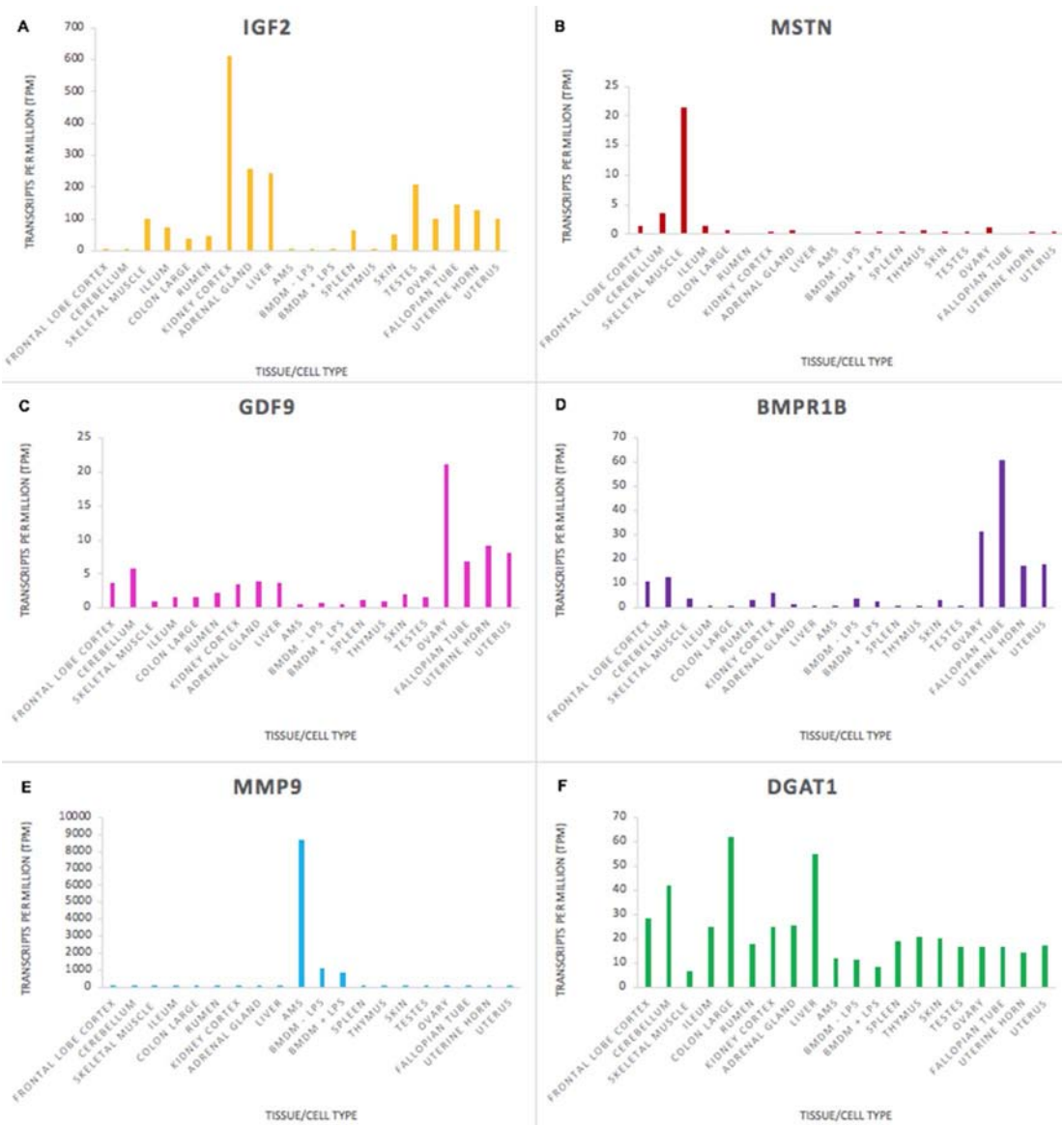


FIGURE 4 | Expression levels (transcripts per million) of genes involved in functional traits in goats to illustrate tissue and cell type or ubiquitous expression patterns in the mini atlas dataset. **(A)** *IGF2* is associated with growth rate; **(B)** *MSTN* is associated with muscle characteristics; **(C)** *GDF9* is associated with ovulation rate; **(D)** *BMPR1* is associated with fecundity; **(E)** *MMP9* is associated with resistance to mastitis; **(F)** *DGAT1* is associated with fat content in goat milk.

and litter size in goats, including *GDF9* and *BMPR1B* (Feng et al., 2011; Shokrollahi and Morammazi, 2018), were highest in the ovary (**Figures 4C, D**). The ovary included here is from a neonatal goat and these results correlate with similar observations in sheep where genes essential for ovarian follicular growth and involved in ovulation rate regulation and fecundity were highly expressed in foetal ovary at 100 days gestation (Clark et al., 2017).

Some genes, particularly those involved in the immune response, had high tissue or cell type specific expression. Matrix metalloproteinase-9 (*MMP9*), which is involved in the inflammatory response and linked to mastitis regulation in goats (Li et al., 2016) was very highly expressed in macrophages, particularly AMs, in comparison with other tissues (**Figure 4E**). Other genes that are important for goat functional traits were fairly ubiquitously expressed. The expression level of Diacylglycerol O-Acyltransferase 1 (*DGAT1*) which is associated with milk fat content in dairy goats (Martin et al., 2017) did not vary hugely across the tissues sampled (**Figure 4F**), although there was slightly higher expression in some tissues (e.g., colon and liver) relative to immune tissues (e.g., thymus and spleen). *DGAT1* encodes a key metabolic enzyme that catalyses the last, and rate-limiting step of triglyceride synthesis, the transformation from a diacylglycerol to a triacylglycerol (Bell and Coleman, 1980). This is an important cellular process undertaken by the majority of cells, explaining its ubiquitous expression pattern. Two exonic mutations in the *DGAT1* gene in dairy goats have been associated with a notable decrease in milk fat content (Martin et al., 2017). Understanding how these, and other variants for functional traits, are expressed can help us to determine how their effect on gene expression and regulation influences the observed phenotypes in goat breeding programmes.

Allele-Specific Expression

Using mapping bias correction for robust positive ASE discovery (Salavati et al., 2019), we were able to profile moderate to extreme allelic imbalance across tissues and cell types, at the gene level, in goats. The raw ASE values for every tissue/cell type are included in **Supplementary Dataset S4**. We first calculated the distribution of heterozygote sites per gene, as a measure of homogeneity of input sites, and found there was no significant difference between the eight individual goats included in the study (**Supplementary Figure S2**).

Several genes exhibited pervasive allelic imbalance (i.e., where the same imbalance in expression is shared across several tissues/cell types) (**Figure 5**). For example, allelic imbalance was observed in the mitochondrial ribosomal protein *MRPL17* in 16 tissues/cell types (except skeletal muscle and rumen). *SERPINH1*, a member of the serpin superfamily, was the only gene in which an imbalance in expression was detected in all tissues/cell types. Allelic imbalance was observed in *COL4A1* in 11 tissues, and was highest in the rumen and skin samples. *COL4A1* has been shown to be involved in the growth and development of the rumen papillae in cattle (Nishihara et al., 2018) and sheep (Bush et al., 2019). The highest levels of allelic imbalance in individual genes were observed in ribosomal protein *RPL10A* in ileum and *SPARC* in liver (**Figure 5**).

The ASE profiles were highly tissue- or cell type-specific, with strong correlations between samples from the same organ system (**Figure 6**). For example, ASE profiles in female reproductive system (ovary, fallopian tube, uterine horn, uterus), GI tract (colon and ileum), and brain (cerebellum and frontal lobe cortex) tissues were highly correlated. The two tissues showing the largest proportion of shared allele-specific expression were the ovary and liver (**Figure 6**). This might reflect transcriptional activity in these tissues in neonatal goats during oogenesis (ovary) and haematopoiesis (liver). Future work could determine if these ASE patterns were observed at other stages of development, or whether they are time-dependant.

The next step of this analysis would be to analyse ASE at the variant (SNV) level. This would allow us to identify variants driving ASE and determine whether they were located within important genes for functional traits. These variants could then be weighted in genomic prediction algorithms for genomic selection, for example. The sequencing depth used for the goat mini-atlas is, however, insufficient for statistically robust analysis at the SNV level. Nevertheless, it does provide a foundation for further analysis of ASE relevant to functional traits using a suitable dataset, ideally from a larger number of individuals (e.g., for aseQTL analysis (Wang et al., 2018)) and at a greater depth.

CONCLUSIONS

We have created a mini-atlas of gene expression for the domestic goat. This expression dataset complements the genetic and genomic resources already available for goat (Tosser-Klopp et al., 2014; Stella et al., 2018; Talenti et al., 2018), and provides a set of functional information to annotate the current reference genome (Bickhart et al., 2017; Worley, 2017). We were able to detect the majority (90%) of the transcriptome from a subset of 17 transcriptionally rich tissues and 3 cell-types representing all the major organ systems, providing proof of concept that this mini-atlas approach is useful for studying gene expression and for functional annotation. Using the mini-atlas dataset, we annotated 15% of the unannotated genes in ARS1. Our dataset was also used by the Ensembl team to create a new gene build for the goat ARS1 reference genome (https://www.ensembl.org/Capra_hircus/Info/Index). One limitation of the mini-atlas is that it included only one biological replicate from a female goat because tissue from female dairy goats is difficult to source. Similarly, the samples used to generate the mini-atlas were all collected from neonatal animals and logistical constraints related to sample collection meant we could not sample immune cells from blood. Future studies could build on the mini-atlas, by including additional biological replicates from females, tissues from multiple developmental stages, and additional types of immune cell (e.g., monocytes, T-cells, and B-cells) to capture further transcriptional complexity.

We have also provided transcriptional profiling of macrophages in goats and a comparative analysis with sheep, which indicated in the cell types and animals investigated in this study transcriptional patterns in the two species were similar. This provides a foundation for further analysis in more tissues and cell types in age-matched animals, and in disease challenge experiments for example.

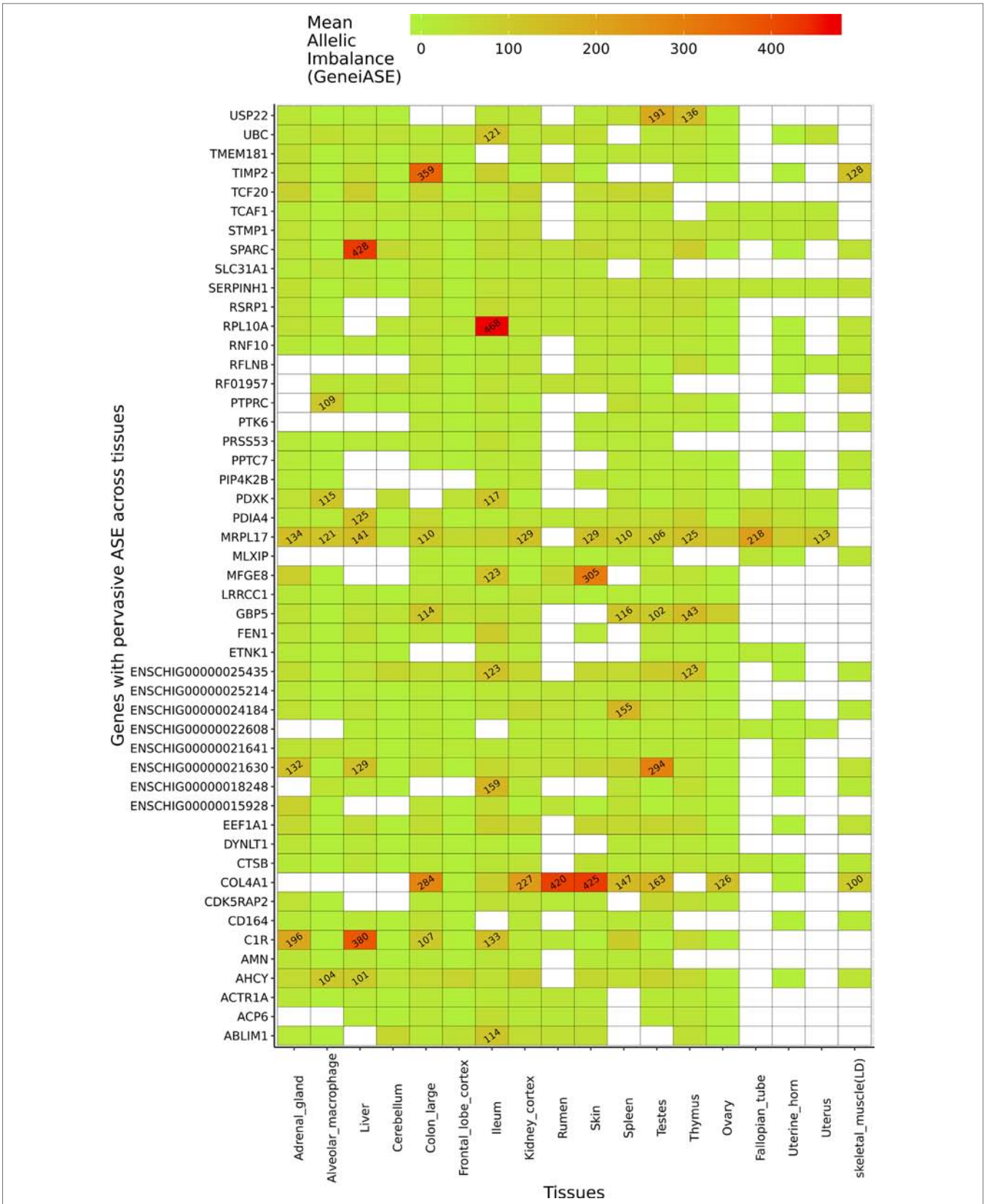


FIGURE 5 | Genes exhibiting the largest mean allelic imbalance (i.e., allele-specific expression averaged across all heterozygote sites within each gene) across 17 tissues and one cell type from the goat mini-atlas dataset visualised as a heatmap (red indicating the highest level of mean allelic imbalance and green the least).

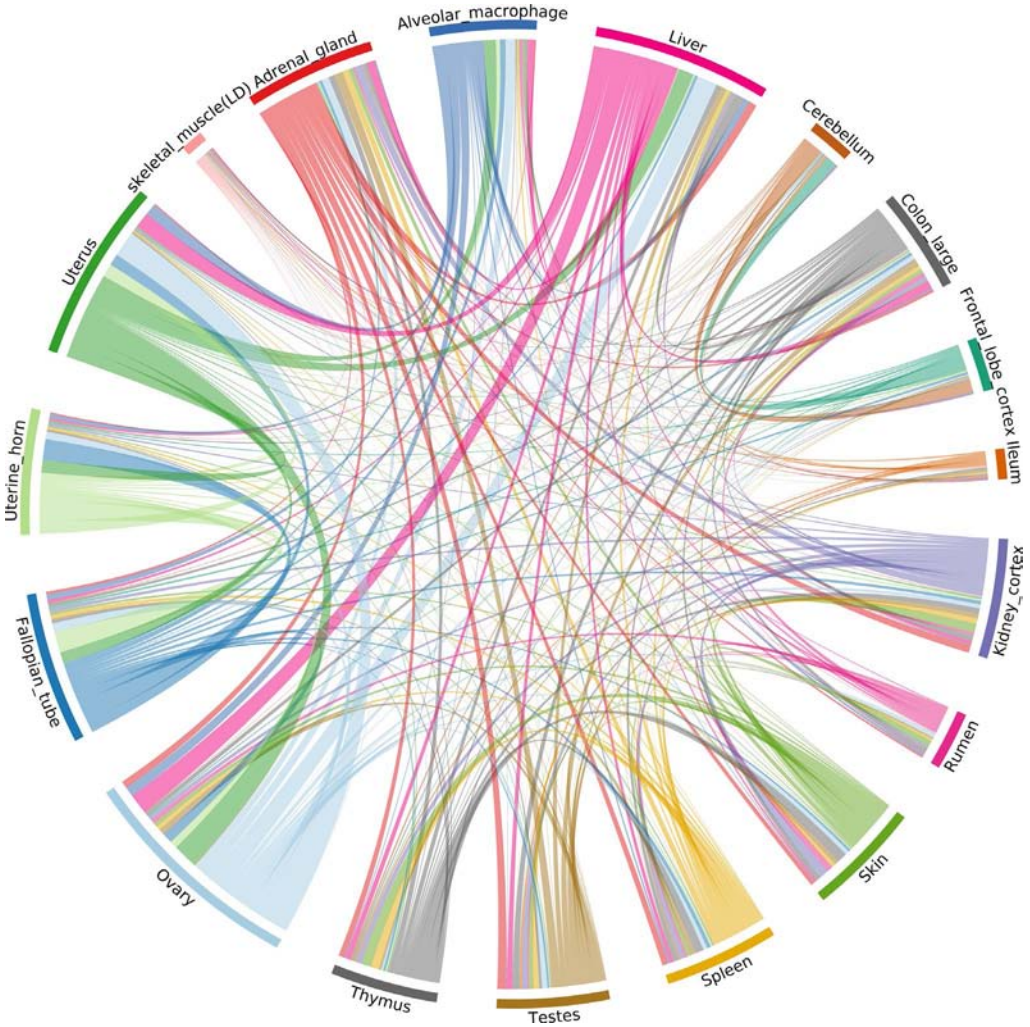


FIGURE 6 | Correlation of allele-specific expression (ASE) profiles shared across tissues/cell types from the goat mini-atlas dataset. Each section represents the genes showing significant allelic imbalance within the tissue. The chords represent the correlation coefficient (CC < 0.85) of ASE profiles shared between the samples (i.e., the proportion of genes showing co-imbalance).

Prior to this study, little was known about the transcription in goat macrophages. While more information is available on goat monocyte derived macrophages (Adeyemo et al., 1997; Taka et al., 2013; Walia et al., 2015), there was previously relatively little knowledge available on the characteristics of goat BMDMs. In addition, few reagents are available for immunological studies in goat, with most studies relying on cross-reactivity with sheep and cattle antibodies (Entrican, 2002; Hope et al., 2012). Recently, a characterisation of goat antibody loci has been published using the new reference genome ARS1 (Schwartz et al., 2018), demonstrating the usefulness of a highly contiguous reference genome with high quality functional annotation for the development of new resources for livestock species. The goat mini-gene expression atlas complements the large gene expression dataset we have generated for sheep and contributes to the genomic resources we are developing for interpretation of the relationship between genotype and phenotype in small ruminants.

DATA AVAILABILITY STATEMENT

We have made the files containing the expression estimates for the goat mini-atlas (**Supplementary Dataset S2** (unaveraged) and **Supplementary Dataset S3** (averaged)) available for download through the University of Edinburgh DataShare portal (https://doi.org/10.7488/ds/2591). Sample metadata for all the tissue and cell samples collected has been deposited in the EBI BioSamples database under project identifier GSB-2131 (https://www.ebi.ac.uk/biosamples/samples/SAMEG330351) according to FAANG metadata and data sharing standards. The raw fastq files for the RNA-Seq libraries are deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena) under the accession number PRJEB23196. The data submission to the ENA includes experimental metadata prepared according to the FAANG Consortium metadata and data sharing standards. The BAM files are also available as analysis files under accession number PRJEB23196 ("BAM file 1"

are mapped to the NCBI version of ARS1 and "BAM file 2" to the Ensembl version). The data from sheep included in this analysis has been published previously and is available *via* (Clark et al., 2017) and under ENA accession number PRJEB19199. Details of all the samples for both goat and sheep are available *via* the FAANG data portal (http://data.faang.org/home). All experimental protocols are available on the FAANG consortium website at http://www.ftp.faang.ebi.ac.uk/ftp/protocols.

ETHICS STATEMENT

The animal study was reviewed and approved by The Roslin Institute, University of Edinburgh's Animal Work and Ethics Review Board (AWERB). All animal work was carried out under the regulations of the Animals (Scientific Procedures) Act 1986.

AUTHOR CONTRIBUTIONS

EC, CM, and DH designed the study. MA, AD, and DH provided guidance on project design, sample collection, and analysis. DH, MA, and AD secured the funding for the project with CM. CM and EC collected the samples with ZL and MM who performed the post mortems. CM performed the RNA extractions. SB performed the bioinformatic analyses. MS performed the analysis of allele-specific expression and assisted CM with the comparative analysis. CM performed the network cluster analysis with EC. CM and EC wrote the manuscript. All authors contributed to editing and approved the final version of the manuscript.

FUNDING

This work was partially supported by a Biotechnology and Biological Sciences Research Council (BBSRC; www.bbsrc.ac.uk) grant BB/L001209/1 ('Functional Annotation of the Sheep Genome') and Institute Strategic Program grants 'Blueprints for Healthy Animals' (BB/P013732/1) and 'Improving Animal Production and Welfare' (BB/P013759/1). The goat RNA-seq data was funded by the Roslin Foundation (www.roslinfoundation.com), which also supported SB. CM was supported by a Newton Fund PhD studentship (www.newtonfund.ac.uk). EC is supported by a University of Edinburgh Chancellor's Fellowship. Edinburgh Genomics is partly supported through core grants from the BBSRC (BB/J004243/1), National Research Council (NERC; www.nationalacademies.org.uk/nrc) (R8/H10/56), and Medical Research Council (MRC; www.mrc.ac.uk) (MR/K001744/1). Open access fees were covered by an RCUK block grant to the University of Edinburgh for article processing charges. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Lindsey Waddell, Anna Raper, Rahki Harne, Rachel Young, Lucas Lefevre, and Lucy Freem for assistance with isolating and characterising BMDMs. Peter Harrison and Jun Fan at the FAANG Data Coordination Centre provided advice on upload of raw data, sample, and experimental metadata to the ENA and BioSamples. This manuscript has been released as a pre-print at BioRxiv (Muriuki et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01080/full#supplementary-material

REFERENCES

Adeyemo, O., Gao, R. J., and Lan, H. C. (1997). Cytokine production in vitro by macrophages of goats with caprine arthritis-encephalitis. *Cell Mol. Biol.* 43 (7), 1031–1037.

Alba-Hurtado, F., and Munoz-Guzman, M. A. (2013). Immune responses associated with resistance to haemonchosis in sheep. *Biomed. Res. Int.* 162158–162169. doi: 10.1155/2013/162158

Alexa, A., and Rahnenfuhrer, J. (2010). *topGO: Enrichment analysis for Gene Ontology* [Online]. Available: http://www.bioconductor.org/packages/release/bioc/html/topGO.html.

Andersson, L., Archibald, A. L., Bottema, C. D., Brauning, R., Burgess, S. C., Burt, D. W., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol* 16 (1), 57. doi: 10.1186/s13059-015-0622-4

Baillie, J. K., Arner, E., Daub, C., De Hoon, M., Itoh, M., Kawaji, H., et al. (2017). Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genet* 13 (3), e1006641. doi: 10.1371/journal.pgen.1006641

Bell, R. M., and Coleman, R. A. (1980). Enzymes of Glycerolipid Synthesis in Eukaryotes. *Annu. Rev. Biochem.* 49 (1), 459–487. doi: 10.1146/annurev.bi.49.070180.002331

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49 (4), 643–650. doi: 10.1038/ng.3802

Bishop, S. C., and Morris, C. A. (2007). Genetics of disease resistance in sheep and goats. *Small Rumin Res* 70 (1), 48–59. doi: 10.1016/j.smallrumres.2007.01.006

Bishop, S. C., and Stear, M. J. (2003). Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections. *Vet Parasitol* 115 (2), 147–166. doi: 10.1016/s0304-4017(03)00204-8

Bishop, S. C., and Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livest Sci* 166, 190–198. doi: 10.1016/j.livsci.2014.04.034

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotech.* 34, 525–527. doi: 10.1038/nbt.3519

Burren, A., Neuditschko, M., Signer-Hasler, H., Frischknecht, M., Reber, I., Menzi, F., et al. (2016). Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Anim. Genet.* 47 (6), 727–739. doi: 10.1111/age.12476

Bush, S. J., Freem, L., MacCallum, A. J., O'Dell, J., Wu, C., Afrasiabi, C., et al. (2018a). Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken. *BMC Genomics* 19 (1), 594–594. doi: 10.1186/s12864-018-4972-7

Bush, S. J., McCulloch, M. E. B., Muriuki, C., Salavati, M., Davis, G. M., Farquhar, I. L., et al. (2019). Comprehensive transcriptional profiling of the gastrointestinal

tract of ruminants from birth to adulthood reveals strong developmental stage specific gene expression. *G3 (Bethesda)* 9 (2), 359. doi: 10.1534/g3.118.200810

Bush, S. J., Muriuki, C., McCulloch, M. E. B., Farquhar, I. L., Clark, E. L., and Hume, D. A. (2018b). Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *GSE* 50 (1), 20. doi: 10.1186/s12711-018-0391-0

Cecchi, F., Russo, C., Iamartino, D., Galiero, A., Turchi, B., Fratini, F., et al. (2017). Identification of candidate genes for paratuberculosis resistance in the native Italian Garfagnina goat breed. *Trop Anim Health Prod* 49 (6), 1135–1142. doi: 10.1007/s11250-017-1306-8

Clark, E. L., Bush, S. J., McCulloch, M. E. B., Farquhar, I. L., Young, R., Lefevre, L., et al. (2017). A high resolution atlas of gene expression in the domestic sheep (Ovis aries). *PLOS Genet* 13 (9), e1006997. doi: 10.1371/journal.pgen.1006997

Dinarello, C. A. (2018). Overview of the IL-1 family in innate inflammation and acquired immunity. *Immunol Rev.* 281 (1), 8–27. doi: 10.1111/imr.12621

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31 (2), 135–141. doi: 10.1038/nbt.2478

Edsgärd, D., Iglesias, M. J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J., et al. (2016). GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci Rep* 6, 21134. doi: 10.1038/srep21134

Entrican, G. (2002). New technologies for studying immune regulation in ruminants. *Vet Immunol and Immunopathol* 87, (3–4), 485–490.

FAANG. (2017). FAANG Data Portal [Online]. Available: http://data.faang.org/home.

Feng, T., Geng, C. X., Lang, X. Z., Chu, M. X., Cao, G. L., Di, R., et al. (2011). Polymorphisms of caprine GDF9 gene and their association with litter size in Jining Grey goats. *Mol Biol Rep* 38 (8), 5189–5197. doi: 10.1007/s11033-010-0669-y

Freeman, T. C., Goldovsky, L., Brosch, M., van Dongen, S., Mazière, P., et al. (2007). Construction, visualisation and clustering of transcription networks from micorarray expression data. *PLoS Comp. Biol.* 3 (10), e206. doi: 10.1371/journal.pcbi.0030206

Freeman, T. C., Ivens, A., Baillie, J. K., Beraldi, D., Barnett, M. W., Dorward, D., et al. (2012). A gene expression atlas of the domestic pig. *BMC Biol.* 10 (1), 90. doi: 10.1186/1741-7007-10-90

Gill, H. S., Altmann, K., Cross, M. L., and Husband, A. J. (2000). Induction of T helper 1- and T helper 2-type immune responses during Haemonchus contortus infection in sheep. *Immunol.* 99 (3), 458–463. doi: 10.1046/j.1365-2567.2000.00974.x

Gossner, A. G., Venturina, V. M., Peers, A., Watkins, C. A., and Hopkins, J. (2012). Expression of sheep interleukin 23 (IL23A, alpha subunit p19) in two distinct gastrointestinal diseases. *Vet Immunol Immunopathol* 150 (1–2), 118–122. doi: 10.1016/j.vetimm.2012.08.004

Harrison, P. W., Fan, J., Richardson, D., Clarke, L., Zerbino, D., Cochrane, G., et al. (2018). FAANG, establishing metadata standards, validation and best practices for the farmed and companion animal community. *Anim. Genet.* 49 (6), 520–526. doi: 10.1111/age.12736

Hope, J. C., Sopp, P., Wattegedera, S., and Entrican, G. (2012). Tools and reagents for caprine immunology. *Small Rumin Res* 103 (1), 23–27. doi: 10.1016/j.smallrumres.2011.10.015

Hume, D. A. (2015). The Many Alternative Faces of Macrophage Activation. *Front. Immunol.* 6, 370–370. doi: 10.3389/fimmu.2015.00370

Ikemizu, T., Kitamura, N., Yamada, J., and Yamashita, T. (1994). Is Lamina Muscularis Mucosae Present in the Ruminal Mucosa of Cattle? *Anat. Histol. Embryol.* 23 (2), 177–186. doi: 10.1111/j.1439-0264.1994.tb00250.x

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Meth.* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Krupp, M., Marquardt, J. U., Sahin, U., Galle, P. R., Castle, J., and Teufel, A. (2012). RNA-Seq Atlas - A reference database for gene expression profiling in normal tissue by next generation sequencing. *Bioinf.* 28 (8), 1184–1185. doi: 10.1093/bioinformatics/bts084

Larroque, H., Rupp, R., Conington, J., Mucha, S., and McEwan, J. (2016). Genomic application in sheep and goat breeding. *Anim. Front.* 6 (1), 39–44. doi: 10.2527/af.2016-0006

Li, H., Zheng, H., Li, L., Shen, X., Zang, W., and Sun, Y. (2016). The Effects of Matrix Metalloproteinase-9 on Dairy Goat Mastitis and Cell Survival of Goat Mammary Epithelial Cells. *PLoS One* 11 (8), e0160989. doi: 10.1371/journal.pone.0160989

Livigni, A., O'Hara, L., Polak, M. E., Angus, T., Wright, D. W., Smith, L. B., et al. (2018). A graphical and computational modeling platform for biological pathways. *Nat. Protoc.* 13, 705. doi: 10.1038/nprot.2017.144

Lopetuso, L. R., Chowdhry, S., and Pizarro, T. T. (2013). Opposing Functions of Classic and Novel IL-1 Family Members in Gut Health and Disease. *Front. Immunol.* 4, 181. doi: 10.3389/fimmu.2013.00181

Martin, P., Palhière, I., Maroteau, C., Bardou, P., Canale-Tabet, K., Sarry, J., et al. (2017). A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Sci Rep* 7 (1), 1872. doi: 10.1038/s41598-017-02052-0

Masahiro, Y., Craig, N. J., Laurie, J. K., and John, D. R. (2006). The sheep and cattle Peyer's patch as a site of B-cell development. *Vet Res* 37 (3), 401–415.

Muriuki, C., Bush, S. J., Salavati, M., McCulloch, M. E. B., Lisowski, Z. M., Agaba, M., et al. (2019). A mini-atlas of gene expression for the domestic goat (*Capra hircus*) reveals transcriptional differences in immune signatures between sheep and goats. *BioRxiv*. doi: 10.1101/711127

Nishihara, K., Kato, D., Suzuki, Y., Kim, D., Nakano, M., Yajima, Y., et al. (2018). Comparative transcriptome analysis of rumen papillae in suckling and weaned Japanese Black calves using RNA sequencing. *J. Anim. Sci.* 96 (6), 2226–2237. doi: 10.1093/jas/skx016

Oliver, S. (2000). Proteomics: Guilt-by-association goes global. *Nat.* 403 (6770), 601–603. doi: 10.1038/35001165

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotech.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, D61–D65. doi: 10.1093/nar/gkl842

Pulina, G., Milán, M. J., Lavín, M. P., Theodoridis, A., Morin, E., Capote, J., et al. (2018). Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. *J. Dairy. Sci.* 101 (8), 6715–6729. doi: 10.3168/jds.2017-14015

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinf.* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Salavati, M., Bush, S. J., Palma-Vera, S., McCulloch, M. E. B., Hume, D. A., and Clark, E. L. (2019). Elimination of reference mapping bias reveals robust immune related allele-specific expression in cross-bred sheep. *Front. Genet.* 10:863. doi: 10.3389/fgene.2019.00863

Schroder, K., Irvine, K. M., Taylor, M. S., Bokil, N. J., Le Cao, K.-A., Masterman, K.-A., et al. (2012). Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. *PNAS* 109 (16), E944–E953. doi: 10.1073/pnas.1110156109

Schwartz, J. C., Philp, R. L., Bickhart, D. M., Smith, T. P. L., and Hammond, J. A. (2018). The antibody loci of the domestic goat (Capra hircus). *Immunogenet.* 70 (5), 317–326. doi: 10.1007/s00251-017-1033-3

Sherman, D. M. (2011). The spread of pathogens through trade in small ruminants and their products. *Rev Sci Tech*, 30 (1), 207–217.

Shokrollahi, B., and Morammazi, S. (2018). Polymorphism of GDF9 and BMPR1B genes and their association with litter size in Markhoz goats. *Repro Domest Anim* 53 (4), 971–978. doi: 10.1111/rda.13196

Stella, A., Nicolazzi, E. L., Van Tassell, C. P., Rothschild, M. F., Colli, L., Rosen, B. D., et al. (2018). AdaptMap: exploring goat diversity and adaptation. *GSE* 50 (1), 61. doi: 10.1186/s12711-018-0427-5

Taka, S., Liandris, E., Gazouli, M., Sotirakoglou, K., Theodoropoulos, G., Bountouri, M., et al. (2013). In vitro expression of the SLC11A1 gene in goat monocyte-derived macrophages challenged with Mycobacterium avium subsp paratuberculosis. *Infect. Genet. Evol.* 17, 8–15. doi: 10.1016/j.meegid.2013.03.033

Talenti, A., Palhière, I., Tortereau, F., Pagnacco, G., Stella, A., Nicolazzi, E. L., et al. (2018). Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. *GSE* 50 (1), 55. doi: 10.1186/s12711-018-0423-9

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., et al. (2014). Design and Characterization of a 52K SNP Chip for Goats. *PLoS One* 9 (1), e86227. doi: 10.1371/journal.pone.0086227

Treanor, B. (2012). B-cell receptor: from resting state to activate. *Immunol.* 136 (1), 21–27. doi: 10.1111/j.1365-2567.2012.03564.x

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J. K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Meth.* 12, 1061. doi: 10.1038/nmeth.3582

Waddell, L. A., Lefevre, L., Bush, S. J., Raper, A., Young, R., Lisowski, Z. M., et al. (2018). ADGRE1 (EMR1, F4/80) Is a Rapidly-Evolving Gene Expressed in Mammalian Monocyte-Macrophages. *Front. Immunol.* 9, 2246.

Walia, V., Kumar, R., and Mitra, A. (2015). Lipopolysaccharide and Concanavalin A Differentially Induce the Expression of Immune Response Genes in Caprine Monocyte Derived Macrophages. *Anim. Biotechnol.* 26 (4), 298–303. doi: 10.1080/10495398.2015.1013112

Wang, M., Hancock, T. P., Chamberlain, A. J., Vander Jagt, C. J., Pryce, J. E., Cocks, B. G., et al. (2018). Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. *BMC Genomics* 19, 395. doi: 10.1186/s12864-018-4800-0

Wang, M., Yu, H., Kim, Y. S., Bidwell, C. A., and Kuang, S. (2012). Myostatin facilitates slow and inhibits fast myosin heavy chain expression during myogenic differentiation. *Biochem. Biophys. Res. Commun.* 426 (1), 83–88. doi: 10.1016/j.bbrc.2012.08.040

Wickham, H., François, R., Henry, L., and Müller, K., (2018). *dplyr: A Grammar of Data Manipulation. R package version 0.7.6*. [Online]. Available: https://CRAN.R-project.org/package=dplyr.

Wickramasinghe, S., Cánovas, A., Rincón, G., and Medrano, J. F. (2014). RNA-Sequencing: A tool to explore new frontiers in animal genetics. *Livestock Sci.* 166, 206–216. doi: 1016/j.livsci.2014.06.015

Worley, K. C. (2017). A golden goat genome. *Nat. Genet.* 49 (4), 485–486. doi: 10.1038/ng.3824

Xiang, R., McNally, J., Rowe, S., Jonker, A., Pinares-Patino, C. S., Oddy, V H., et al. (2016a). Gene network analysis identifies rumen epithelial cell proliferation, differentiation and metabolic pathways perturbed by diet and correlated with methane production. *Sci Rep* 6, 39022. doi: 10.1038/srep39022

Xiang, R., Oddy, V. H., Archibald, A. L., Vercoe, P. E., and Dalrymple, B. P. (2016b). Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues. *PeerJ* 4, e1762. doi: 10.7717/peerj.1762

Young, R., Bush, S. J., Lefevre, L., McCulloch, M. E. B., Lisowski, Z. M., Muriuki, C., et al. (2018). Species-Specific Transcriptional Regulation of Genes Involved in Nitric Oxide Production and Arginine Metabolism in Macrophages. *ImmunoHorizons* 2 (1), 27.

Yu, B., Lu, R., Yuan, Y., Zhang, T., Song, S., Qi, Z., et al. (2016). Efficient TALEN-mediated myostatin gene editing in goats. *BMC Dev. Biol.* 16 (1), 26. doi: 10.1186/s12861-016-0126-9

Zerbino, D. R., Achuthan, P., Akanni, W., AmodeBarrell, M. R.D., Bhai, J., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46 (D1), D754–D761. doi: 10.1093/nar/gkx1098

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Muriuki, Bush, Salavati, McCulloch, Lisowski, Agaba, Djikeng, Hume and Clark. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 November 2019
doi: 10.3389/fgene.2019.01032

Analysis of the Progeny of Sibling Matings Reveals Regulatory Variation Impacting the Transcriptome of Immune Cells in Commercial Chickens

*Lucy Freem[1†], Kim M. Summers[2†], Almas A. Gheyas[1], Androniki Psifidi[3], Kay Boulton[1], Amanda MacCallum[1], Rakhi Harne[1], Jenny O'Dell[1], Stephen J. Bush[4] and David A. Hume[2]**

[1] The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, [2] Mater Research Institute-University of Queensland, Translational Research Institute, Woolloongabba, QLD, Australia, [3] Department of Clinical Sciences and Services, Royal Veterinary College, University of London, London, United Kingdom, [4] Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
Yuanning Li,
Yale University, United States
Brittney Keel,
United States Department of Agriculture, United States

***Correspondence:**
David Hume
David.Hume@uq.edu.au

†These authors contributed equally to the study

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 03 July 2019
Accepted: 25 September 2019
Published: 14 November 2019

Citation:
Freem L, Summers KM, Gheyas AA, Psifidi A, Boulton K, MacCallum A, Harne R, O'Dell J, Bush SJ and Hume DA (2019) Analysis of the Progeny of Sibling Matings Reveals Regulatory Variation Impacting the Transcriptome of Immune Cells in Commercial Chickens. Front. Genet. 10:1032. doi: 10.3389/fgene.2019.01032

There is increasing recognition that the underlying genetic variation contributing to complex traits influences transcriptional regulation and can be detected at a population level as expression quantitative trait loci. At the level of an individual, allelic variation in transcriptional regulation of individual genes can be detected by measuring allele-specific expression in RNAseq data. We reasoned that extreme variants in gene expression could be identified by analysis of inbred progeny with shared grandparents. Commercial chickens have been intensively selected for production traits. Selection is associated with large blocks of linkage disequilibrium with considerable potential for co-selection of closely linked "hitch-hiker alleles" affecting traits unrelated to the feature being selected, such as immune function, with potential impact on the productivity and welfare of the animals. To test this hypothesis that there is extreme allelic variation in immune-associated genes we sequenced a founder population of commercial broiler and layer birds. These birds clearly segregated genetically based upon breed type. Each genome contained numerous candidate null mutations, protein-coding variants predicted to be deleterious and extensive non-coding polymorphism. We mated selected broiler-layer pairs then generated cohorts of F2 birds by sibling mating of the F1 generation. Despite the predicted prevalence of deleterious coding variation in the genomic sequence of the founders, clear detrimental impacts of inbreeding on survival and post-hatch development were detected in only one F2 sibship of 15. There was no effect on circulating leukocyte populations in hatchlings. In selected F2 sibships we performed RNAseq analysis of the spleen and isolated bone marrow-derived macrophages (with and without lipopolysaccharide stimulation). The results confirm the predicted emergence of very large differences in expression of individual genes and sets of genes. Network analysis of the results identified clusters of co-expressed genes that vary between individuals and suggested the existence of trans-acting variation in the expression in macrophages of the interferon response factor family that distinguishes the parental broiler and layer birds and influences the global response

to lipopolysaccharide. This study shows that the impact of inbreeding on immune cell gene expression can be substantial at the transcriptional level, and potentially opens a route to accelerate selection using specific alleles known to be associated with desirable expression levels.

Keywords: chicken, genome, inbreeding, allele-specific, transcriptome, macrophage

INTRODUCTION

A large proportion of causal genetic variation implicated in complex traits in humans is associated with regulatory variants that impact on the level of gene expression (Zhu et al., 2016). Gene expression is itself a complex trait, controlled by both *cis*-acting and *trans*-acting (epistatic) variants and interactions with environment (GTEx project, 2017). Gene expression can therefore provide an intermediate phenotype in analysis of complex traits, an approach that has been termed genetical genomics [(Johnsson et al., 2018) and references therein]. At a population level, expression quantitative trait loci (eQTL) studies of individual cells or tissues can reveal associations between single nucleotide variants (SNVs) and the amount of each mRNA transcribed from the genome (GTEx project, 2017). At the level of an individual, provided there are expressed SNVs, RNA sequencing enables the identification of regulatory variation within a locus, based upon the relative expression of the two alleles (so-called allele-specific expression, ASE) (Pastinen, 2010).

Modern western broiler and layer chickens have been divergently selected for meat and egg production respectively, increasingly using genomic selection with dense genotyping chips (Kranis et al., 2013; Gheyas et al., 2015). This intense selection has generated selective sweeps around regions associated with production traits, for example genes linked to appetite, growth, metabolic regulation and carcase traits in broilers and egg-production in layers (Qanbari et al., 2019). Analysis of linkage disequilibrium in commercial broiler, white egg and brown egg layers revealed highly divergent patterns between selected populations, significant inbreeding coefficients and, on average, much larger average LD blocks than in human populations (Pengelly et al., 2016). The selection for production traits could potentially impact inadvertently (through co-selection) or directly on the immune system which makes an important contribution to fitness in production animals. In the current study we take a novel approach to identifying the potential immunological consequences of trait selection in commercial chickens.

Macrophages are an essential component of the innate immune system. In mammals, the proliferation and differentiation of macrophages depends upon signaling through the macrophage colony-stimulating factor receptor (CSF1R) *via* two ligands, CSF1 and interleukin 34 (IL34). This system is functionally conserved in birds (Garceau et al., 2010). Recombinant CSF1 can be used to generate pure populations of macrophages *in vitro* from bone marrow progenitors (Garceau et al., 2010). We used this system to demonstrate that genes on the Z chromosome in birds are generally not fully dosage compensated in male (ZZ) versus female (ZW) birds. We showed also that the presence of the interferon genes on the Z chromosome impacts on the relative response of male and female macrophages to bacterial lipopolysaccharide (LPS) (Garcia-Morales et al., 2015). To analyze chicken macrophage biology *in vivo* we have produced *CSF1R* reporter transgenic lines on a conventional layer genetic background (Balic et al., 2014; Garceau et al., 2015).

There is a strong signature of selection over the *CSF1R* locus in commercial broilers (Stainton et al., 2017). Analysis of the genomic sequence data for commercial birds (Gheyas et al., 2015) revealed high prevalence non-synonymous protein-coding variants in *CSF1R* that are unique to either broilers or layers (Hume et al., 2019). In support of the possibility that this variation is functionally significant, mutations in either *Csf1r* or *Csf1* in both mice and rats produce severe post-natal growth retardation (Dai et al., 2002; Pridans et al., 2018). Such variation could obviously also impact on innate immune function. Chicken meat and egg production at scale generally involves housing in well-controlled environments and infection control with vaccines and/or prophylactic antibiotics. These production systems may mask the impact of selection on immune-related traits. Increasingly, the efficacy of vaccines is challenged by pathogen evolution and antibiotic use is now largely prohibited. There has therefore been a renewed interest in breeding for disease resistance and in the identification of markers of disease severity and prognosis. One novel strategy for improving disease resistance is based upon selective breeding of birds that display high levels of inducible pro-inflammatory cytokines (IL6 or the CXCL chemokines) in response to bacterial stimuli (Swaggerty et al., 2008; Swaggerty et al., 2016; Swaggerty et al., 2017; Swaggerty et al., 2019).

Most potential regulatory and protein-coding variants of large effect in commercial birds are masked because of the breeding pyramid approach used. Independent pedigree lines are intensively selected for specific traits and then crossed to maximize heterozygosity in the production animals which may contain genetic contributions from as many as eight heavily-selected founder lines. One presumption in such breeding pyramids is that maximal heterozygosity conceals potentially deleterious alleles; leading to hybrid vigor or heterosis. The reciprocal of heterosis is the well-documented phenomenon of inbreeding depression (Charlesworth and Willis, 2009; Chen, 2013). The molecular basis for both phenomena has been studied more extensively in plants than in animals. At least some of the variation underlying heterosis is regulatory and can be detected at the level of mRNA expression of individual genes, where the level of expression in an F1 hybrid commonly lies at the midpoint of expression of parental lines. RNAseq analysis has been used to address this prediction in a defined cross of two inbred chicken lines, in which gene expression was compared in the brain and liver of the parents and embryonic

F1 birds. This approach provided strong evidence of frequent allelic imbalance in embryonic brain and liver. In the large majority of cases the combined expression of individual transcripts from the two parental alleles in the F1 animals was essentially additive (Zhuo et al., 2017; Zhuo et al., 2019). A small subset of transcripts showed evidence of dominance or over-dominance in expression level; these were identified as candidate *trans*-regulators contributing to heterosis. The key conclusion from these studies is that the *cis*-acting regulation in the parent lines was demonstrable and heritable in the F1 progeny (as allele-specific expression).

The identification of functional allelic variants based upon RNAseq depends upon the presence of informative expressed SNVs in each individual. An alternative approach that is practical in birds, which are multiparous, is to brother-sister mate F1 progeny from a defined parental cross to generate an F2 population in which the grandparental allelic variants at each locus will be homozygous in a subset of birds. Based upon the results from the inbred cross (Zhuo et al., 2017; Zhuo et al., 2019) such a breeding strategy should expose high and low expression alleles that are masked in heterozygotes. Because of their genetic divergence, broiler-layer intercrosses have been used extensively in QTL mapping of production traits [e.g. (Campos et al., 2009; Hocking et al., 2012; Podisi et al., 2013)]. Aside from their regulated expression of genes encoding immunological functions, macrophages also express a very large proportion of the entire transcriptome at detectable levels (Bush et al., 2018). Given their complex transcriptome and the regulatory functions of macrophages in growth and development, it is conceivable genetic variants that control expression of genes involved in the production traits that distinguish broilers and layers also regulate their expression in macrophages or other cells of the immune system. Therefore, to explore these concepts, we have generated a series of families of F2 individuals from sibling matings derived from a cross between commercial broilers and our *CSF1R*-mApple transgenic line (Balic et al., 2014) which is maintained on an outbred layer background and which expresses the mApple reporter in cells of macrophage lineage. Our analysis of expression variance in immune cells in these inbred birds supports the existence of strong allele-specific expression variants in the parental commercial birds.

MATERIALS AND METHODS

Ethical Approval

All animal work including breeding and care was conducted in accordance with guidelines of the Roslin Institute and the University of Edinburgh and carried out under the regulations of the Animals (Scientific Procedures) Act 1986 under Home Office project license PPL 60/4420. Approval was obtained from the Protocols and Ethics Committees of the Roslin Institute and the University of Edinburgh.

Animals

Commercial Ross 308 broilers as founders were obtained as hatchlings from PD Hook (Hatcheries) Ltd, Cote, Brampton, Oxfordshire, UK. Founders from the *CSF1R*-mApple reporter transgenic layer line on an ISA-Brown genetic background (Balic et al., 2014) were produced in The Roslin Institute. All birds were bred and housed in approved facilities within the National Avian Research Facility at The Roslin Institute.

Cell Culture and mRNA Isolation

Bone marrow cells from the femurs of adult or hatchling birds were harvested and cultured for 7 days in recombinant chicken CSF1 to generate a population of bone marrow-derived macrophages (BMDM) as described previously (Garceau et al., 2015; Garcia-Morales et al., 2015; Bush et al., 2018). The macrophages were detached from the plates and re-seeded in 6 well plates with CSF1, with or without 100 ng/ml of LPS, for 24h prior to harvest and purification of mRNA (Garcia-Morales et al., 2015). Spleens were obtained immediately after euthanasia, snap frozen in entirety and stored in RNA-later at −80°C until used for RNA extraction.

Genetic Analysis

Whole genome sequencing of DNA from the set of founder broiler and transgenic layer lines was performed by Edinburgh Genomics, University of Edinburgh, UK, using the Illumina HiSeqX platform. Sample specific libraries were created from genomic DNA using Illumina SeqLab specific TruSeq Nano High Throughput library preparation kits in conjunction with the Hamilton MicroLab STAR and Clarity LIMS X Edition. The gDNA samples were normalized to the concentration and volume required for the Illumina TruSeq Nano library preparation kits, then sheared to a 450 bp mean insert size. The inserts were ligated with blunt ended, A-tailed, size selected, TruSeq adapters and enriched using eight cycles of PCR amplification. The libraries were normalized, denatured, and pooled in eights for clustering and sequencing. The number of paired reads varied from 177,300,445 to 365,616,593, with an average of 265 million paired reads per sample equating to >30× coverage. The read length varied from 35 bases to 150 bases, with modal read length being 150 bases. Sequence quality was checked with FASTQC (v0.11.8) package. Poor quality bases from the ends of reads were trimmed with Trimmomatic v3.5 software. Reads were trimmed with criteria:

- Remove leading low quality (Phred <20) or N bases (LEADING:20).
- Scan the read with a 10-base wide sliding window, cutting when the average quality per base within a window drops below 20 (SLIDINGWINDOW:10:20).
- After the above operations, drop reads less than 50 bp long (MINLEN:50).

The trimming step retained 54% to 85% of the reads and the trimmed read lengths varied between 50 to 150 bp, with the vast majority of reads between 140–150 bp long. This resulted in overall coverage ranging from 30× to 70×.

Sequence reads from each sample were mapped against chicken reference genome (Galgal5.0) using BWA (v 0.7.8)

with BWA_MEM algorithm. The resultant bam files were further processed to mark duplicate reads using Picard tools (v2.1.1) followed by indel realignment using GATK (v 3.7.0). In order to make sure that no contamination or mislabeling occurred during the sample/data processing steps, we checked the realigned bam files for the presence and absence of HIV1 (GAGAGAGATGGGTGCGAGAG) and HIV2 (GCTGTGCGGTGGTCTTACTT) primer sequences that flank the MacApple transgene. As expected, these primer sequences were present only in the transgenic layer birds while absent in the non-transgenic broiler samples.

Variant calling was initially performed with SAMtools mpileup (v 1.1) in conjunction with bcftools with minimum base and mapping qualities set as 20. The variants were annotated against the genomic features annotated by NCBI using package snpEff (v 4.2). Nonsynonymous SNPs were further predicted for their effects on protein sequence changes using the SIFT algorithm in Variant Effect Predictor (VEP). All the variants were also checked for their overlap with evolutionary constraint elements detected from multiple alignments of 49 bird species (https://pag.confex.com/pag/xxiv/webprogram/Paper21473.html) using bedtools (v 2.22.1).

For the selection of F1 breeding pairs, the genotyping of variants within selected genes (*IL10, CSF1R, IL12B, IL34*) used a custom KASP™ (competitive allele-specific PCR) analysis (Biosearch Technologies, Teddington, UK). SNPs from these genes were chosen based on contrasting genotypes and allele frequency in broiler and layers, and annotation results to include potentially functional variants.

Gene Expression Analysis

Library preparation for RNA sequencing (RNAseq) was also performed by Edinburgh Genomics using the Illumina TruSeq mRNA (poly-A selected) library preparation protocol. mRNA was sequenced at a depth of >40 million strand-specific 75 bp paired end reads per sample, using an Illumina HiSeq 4000. Expression was quantified using the high speed quantification tool Kallisto v0.43.1 (Bray et al., 2016) following procedures detailed previously (Bush et al., 2017; Bush et al., 2018). Kallisto quantifies expression at the transcript level by building an index of k-mers from a set of reference transcripts and then mapping the RNA-seq reads to it, matching k-mers generated from the reads with the k-mers present in the index. Transcript-level estimates (transcripts per million, TPM) are then summarised to the gene level. The current analysis used the revised chicken reference transcriptome defined previously (Bush et al., 2018).

Network Analysis

Network analysis was performed using Graphia (Kajeka Ltd, Edinburgh, UK). This software builds a correlation matrix based on gene expression patterns, either for sample-to-sample or gene-to-gene comparisons. A network graph was constructed for all relationships above a threshold Pearson correlation coefficient (as detailed in *Results*), connecting nodes (genes) by edges (correlations between nodes above the threshold). Clustering of nodes within the network was performed for the gene-to-gene analysis using the Markov clustering (MCL) algorithm at an inflation value of 1.7 to decrease granularity resulting from the similarity of samples all from the same cell type. Functional annotation of clusters of genes used DAVID 6.8 (https://david.ncifcrf.gov/home.jsp).

Analysis of Gene Expression in a Low Fitness Family

The expression level of genes in the BMDM samples was averaged over all samples in one family (Family H; N = 6) in which there was poor hatch rate and low body weight and all other samples (N = 22 untreated and N = 25 LPS treated). All genes where the value of at least one of these averages was greater than 1 were included in the analysis. The ratio of the average in Family H to the average in other samples was calculated. Gene symbols for all samples with a ratio of >1.5 or less than 0.67 (i.e. with at least a 1.5-fold difference) were identified. To identify possible pathways involved in the low fitness, the gene sets were analyzed using the gene ontology analysis software DAVID (see above). and separately in GATHER (https://changlab.uth.tmc.edu/gather/). The analysis was performed for control samples and samples treated with LPS separately.

RESULTS AND DISCUSSION

Genomic Sequencing

We completed whole genome sequencing of a total of 10 Ross 308 commercial broilers (five of each sex) and 10 *CSF1R*-mApple transgenic layer birds (five of each sex). **Figure 1A** outlines the analysis and annotation pipeline. Initial variant discovery (with minimum base and mapping quality at 20) indicated a transitions to transversions ratio relative to the reference genome slightly over 2:1 for all samples, which is within normal range. The commercial broiler sequences have a slightly higher level of heterozygosity and much higher numbers of singletons than layers as expected given their derivation from multiple pedigree lines. **Figure 1B** shows a principal component analysis (PCA) of these birds based upon a 13.4 M filtered SNV panel derived from the variant call files for the set of birds sequenced (filtration criteria: min. variant quality 30, min. genotype quality 15, and max. rate of missing genotype 20%). The first principal component clearly separates the broilers from the transgenic layers, whilst the second identifies genetic diversity in the layer population. For reasons that will become evident in subsequent analysis described below, we also generated a PCA based solely upon SNVs detected within loci encoding the members of the interferon-responsive factor (IRF) family, which may produce potential trans-acting transcriptional regulators of inducible gene expression in macrophages. **Figure 1C** shows that the first principal component again separates broilers and layers but the second highlights extensive diversity in both populations.

The main purpose of the whole genome analysis was to identify candidate null protein-coding variants in each of the founder birds and also to determine whether there are prevalent broiler and layer-enriched polymorphic haplotypes in our founders that could be deliberately driven to homozygosity by

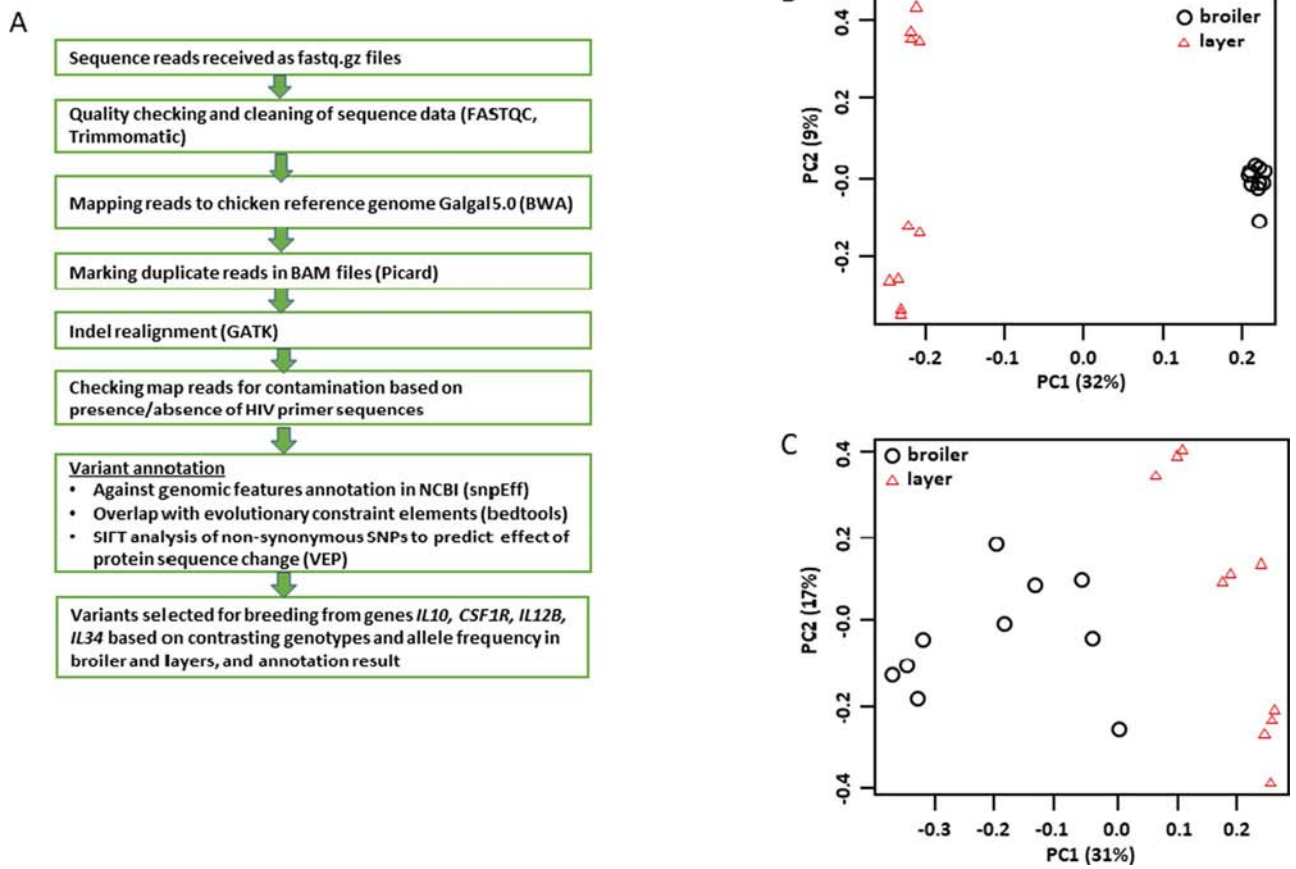


FIGURE 1 | Analysis of genomic sequences of founder layer and broiler birds. Genomic DNA sequences obtained from the 27 commercial broiler and transgenic layer birds used as founders in the inbreeding experiments were analyzed as shown in Panel **(A)**. Panel **(B)** shows a PCA based upon a 13.4M filtered SNV panel derived from the variant call files for the set of birds sequenced. Note the first principle component (PC) separates the layers (left) and the broilers (right). The second PC identifies substantial variation within the two populations with the greatest variation in the layers. Panel **(C)** shows a PCA based solely upon SNVs identified within the members of the IRF gene family (See **Table S6**). Note that PC1 again separates broilers and layers, whereas PC2 identifies substantial variation within breed type.

brother-sister mating. A comparative study of a similar-sized cohort of broiler and layer lines (14 of each) from Brazilian breeds (Boschiero et al., 2018) identified >500 stop gain and >7000 coding variants. Consistent with the genetic diversity identified in this earlier study, the analysis identified a total of 14.1M SNVs in our limited population. The vast majority were non-coding and heterozygous. Within this large set, we identified those predicted to affect exonic sequences, and identified either HIGH impact (i.e. stop gained/lost, start lost, and splice acceptor/donor variants) or nonsynonymous deleterious variants (based on SIFT analysis). We identified 979 SNVs predicted to be HIGH impact and a further 10,872 predicted nonsynonymous coding variants that were annotated as deleterious by SIFT. Often one SNV was associated with more than one predicted impact so it was difficult to place it in a unique group. Amongst the high impact categories that likely produce a complete loss-of-function there were 20 splice_acceptor_variants; 26 splice_donor_variants; 282 start_lost (7 in splice_regions); 628 stop_gained (15 in splice regions); and 142 loss-of-stop codon variants and 1 combined loss-of-start/loss-of-stop. These numbers are approximately consistent with reports of the prevalence of null mutations in chickens based upon much larger cohorts of disparate commercial bird populations (Rubin et al., 2010; Qanbari et al., 2019). Those published studies indicate that few loss-of-function variants are associated with selective sweeps and therefore that such variants are not commonly the direct subject of breeding selection. Indeed, overall amino acid altering mutations were significantly less prevalent in domestic chickens than in their wild red jungle fowl ancestors (Rubin et al., 2010; Qanbari et al., 2019). Nevertheless, there remains very substantial coding sequence polymorphism and potential null mutations in the founder birds that could affect immune and other functions.

Based upon the genome sequencing, we were also able to identify prominent SNVs associated with specific genes of interest that might generate global differences in monocyte-macrophage numbers and/or activation state. **Table 1** summarizes the extensive variation that we identified in the *CSF1R*, *IL34* and *IL10* loci. In addition to *CSF1R* and *IL34* (which could control macrophage differentiation) we focused on *IL10* because of our recent data indicating substantial variance

TABLE 1 | DNA sequence variants in the vicinity of candidate genes of interest detected in 27 founder broiler and layer birds.

Gene	Coordinates in Gal5	No. of Variants	No. of Indels	No. of variants with quality >30	No. HIGH MODERATE effect variants	No. of variants overlapped with constrained elements
CSF1R	Chr 13: 13275507 to 13292566	457	40	453	2	38
CSF1	Chr 26: 1276201 to 1282875	99	21	85	4	0
IL10	Chr 26: 2562275 to 2564509	39	3	38	3	5
IL34	Chr 11: 1776702 to 1781094	98	14	95	1	4

High impact variants include stop gain/loss and splice donor and acceptors. Moderate impact variants includes nonsynonymous SNVs. Constrained elements are evolutionary conserved regions of the chicken genome detected by comparative analysis of 48 bird genomes using GERP++ package (https://pag.confex.com/pag/xxiv/webprogram/Paper21473.html)

in IL10 production amongst commercial broilers responding to Eimeria parasite challenge (Boulton et al., 2018a; Boulton et al., 2018b) and evidence that endogenous IL10 exerts a feedback regulatory effect of macrophages responding to LPS (Wu et al., 2016). As discussed above, the focus on *CSF1R* and its two ligands was based on their possible roles in growth and selection in broilers. The *CSF1* locus was poorly annotated on the chicken genome at the time breeding decisions were made but it is now clear that this locus lies within 2Mb of *IL10* on chromosome 26. Subsequent analysis on a large population of commercial broilers revealed significant LD between *IL10* and *CSF1* and also very limited heterozygosity at the *CSF1* locus itself (Psifidi A, unpublished). From amongst the candidate SNVs detected in *CSF1R* and *IL34* we did not identify protein-coding variants with predicted large effect, but we did identify SNV markers that were strongly-enriched in either the broiler or layer parents and might potentially be linked to expression variants. In the case of IL34, at position 1780913 on Chr 11, neither broilers, nor layers had the reference allele identified in Red Jungle Fowl, and the two were almost fixed for different alternative alleles (C > T and C > G respectively). Accordingly, these variants were used as markers for inbreeding. Within the *IL10* locus we identified a non-synonymous coding variant (p.Cys4Gly) that was more prevalent in the broiler parents.

Each of the founder birds also carried multiple candidate mutations of large effect. **Table 2** summarizes selected examples of high-confidence protein-coding variants in immune-related genes that in most cases were detected in more than one founder as a heterozygote but were not detected as homozygotes. Amongst these candidates, *TNFRSF10B* and *IL12B* were considered possible regulators that might generate an immune-related phenotype if homozygous for loss-of-function alleles. *TNFRSF10* encodes a protein variously known as TRAIL receptor 2, DR5, and Killer. It is involved in triggering of apoptosis. In humans, there are several *TNFRSF10* genes, but in mice, there is a single *TNFRSF10B* gene and null mutation leads to alterations in radiation-induced cell death (Finnberg et al., 2005). The chicken genome (Galgal5) also has a single *TNFRSF10B* gene.

Assuming simple Mendelian inheritance and random mating, around 1/16 birds in the F2 generation will be homozygous for any one of the null mutations present in a heterozygous state in either of its grandparents. In addition to the variants we selected deliberately each F2 bird is likely to be homozygous for a unique subset of the null alleles present in the founder lines. Where the grandparents are homozygous for breed-enriched coding and non-coding allelic variants, all F1 matings are between heterozygotes and 25% of the F2 progeny should be homozygous for any variant if there is no impact on viability.

Primary Phenotype of F2 Birds

In total, we generated 15 families of F2 progeny from brother-sister matings of F1 birds selected by genotyping for heterozygosity for alleles of interest derived from their founder parents. These birds were also selected to be positive for the *CSF1R*-mApple transgene so that the majority of their offspring could be assessed for transgene expression as a marker. For logistical reasons analysis

TABLE 2 | Sequence variants affecting genes expressed in immune cells with predicted HIGH effect on protein coding regions.

Gene	Chr	Pos	Ref	Alt	qual	type	Hom alt	Het	Hom ref
TAPBPL	1	76880284	G	T	999	Splicing	0	9	10
HHLA2	1	87500359	CAGAG	CAGAGAAGAG	197	INDEL frameshift	0	1	18
IL18RAP	1	1.34E+08	TTTTTTTTTTTTTG	T	426	INDEL frameshift	0	9	10
IL8L1	4	51270553	T	A	458	Exonic	0	2	17
TNFSF10	9	19373730	CCTCTC	CCTC	153	INDEL frameshift	0	9	10
TNFRSF10B	22	1281321	C	T	214	Stop gain	0	1	18
IL12B	13	8133187	G	A	999	Exonic	0	5	14
IRF1	13	16983217	GATCCTGTGCTGTGCT	G,GGTGCTGTGCT	960	INDEL frameshift	0	2	9
HMHA1	28	2956121	C	CCA	201	INDEL frameshift	0	2	16
CD69	1	5876	CCAGACAGA	CCAGA	115	INDEL frameshift	0	3	16
CD69	1	5915	GC	G	41	INDEL frameshift	0	3	16
TGFB1	32	27061	CTGGGG	C	416	INDEL	0	2	17
TGFB1	32	28627	CT	C	77	INDEL frameshift	0	7	8
TGFB1	32	28810	CAGCAG	C,CCAG	999	INDEL frameshift	0	14	1

of F2 phenotypes was carried out exclusively in the immediate 48h post-hatch period. All of the F2 birds were weighed at hatch. **Figure 2** shows box and whisker plots for the distribution of body weights in each F2 family. As expected given the nature of the founder lines, one of which (the broilers) has been selected for rapid body weight gain, there was substantial variation in body weight amongst individual F2 birds even at hatch. In some families there were significant outliers but in only one family, family H, was there a significant reduction in average body weight compared to all other families. In that family, three birds died at or just before hatching with substantial abnormalities including exposed brain, blood in thoracic cavity, muscles and liver, underdeveloped cartilage, and abnormal yolk sac and yolk. The overall hatch rate was also substantially reduced (not shown). For any individual family, 1/4 progeny will be homozygotes for any loss-of-function mutations present in both F1 parents and the number of birds assessed in these families is sufficient to identify outliers or fail-to-hatch numbers. Accordingly, the data indicate that few of the loss-of-function mutations identified in WGS lead to compromised development when bred to homozygosity.

Monocyte Count

As a preliminary screen for the possible impact of *CSF1*, *IL34*, or *CSF1R* variants we screened all of the F2 progeny by flow cytometry for the number of blood monocytes (co-expressing the surface marker KUL01) and the prevalence of cells expressing the *CSF1R*-mApple reporter gene. The latter percentage is higher because the reporter gene is also expressed in heterophils, albeit at a 10-fold lower level than in monocytes (Balic et al., 2014). When the data from all F2 hatchlings was merged the two parameters were consistent between birds, with no clear outliers, independent of sex (**Figure 3A**) and did not change when measured on day 1 or day 2 post-hatch (**Figure 3B**). With the available sample sizes, we did not detect a significant effect of homozygosity for the parental *CSF1R* or *IL34* allelic variants. Marginal impacts of heterozygous variation in *TNFRSF10B* and homozygosity for *IL10* variation were insignificant when corrected for the number of comparisons. Overall, this analysis did not provide evidence for variants of large effect in the selected genes that distinguish between broilers and layers and which impact upon the blood heterophil or monocyte count in chickens.

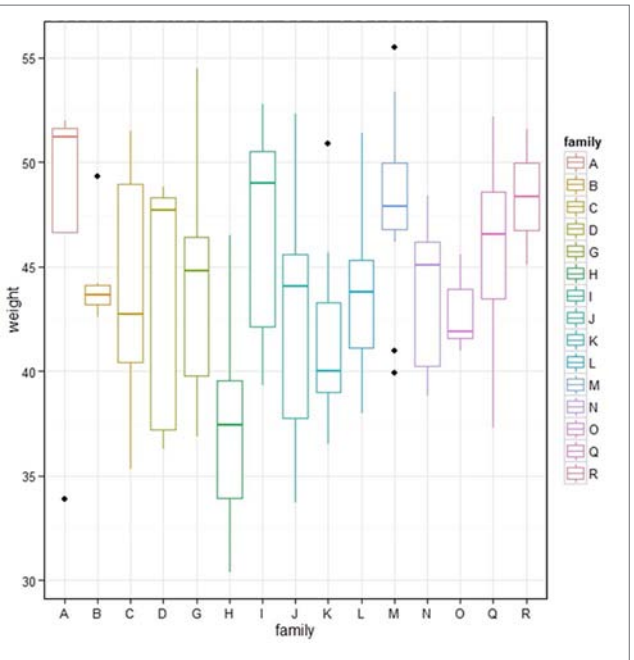


FIGURE 2 | Post hatch body weight in F2 progeny of F1 sibling matings. The figure shows the range of body weights of hatchling birds in each of 15 families derived from sibling mating of the progeny of an F1 broiler-layer cross. Dots identify outliers.

Analysis of Variation of Gene Expression in Spleen of F2 Birds

To screen for variants that influence gene expression in immune cells, we first performed RNAseq on whole spleens from 18 hatchlings derived from F2 families A (1; IL34), C (6; IL12B), G (2; IL10), L (2; IL34), M (3; IL34), Q (3; IL10), R (1; IL10) that segregated the variant alleles for *IL34, IL12B*, and *IL10.* We chose hatchlings to avoid variation that might arise from exposure to infectious agents. Spleen is a mixture of myeloid, lymphoid and other hematopoietic cell populations, each of which has a gene expression signature that can be detected within the total mRNA pool. We profiled males and females from multiple families and individual birds were specifically genotyped for the allelic variants at the *IL34*, *IL12B*, and *IL10* loci to assess whether these variants were associated with any specific pattern of gene expression. SNVs in *IL12B* and *IL10* have been associated with immune-response traits in a large pedigree derived from a white egg/brown egg layer cross (Biscarini et al., 2010) and we considered the possibility that pleiotropic impacts of such genetic variation might manifest in changes in the spleen. In the case of IL34, we reasoned that variation might be associated with monocyte-macrophage number, and hence the relative abundance of macrophage-specific transcripts. For the purpose of the analysis, we removed transcripts that had a maximum expression <10 TPM (see *Methods*). The complete data set from this analysis is provided in **Table S1**.

One consideration in any gene expression analysis in chicken is the difference between males and females. Females are the heterogametic sex and have only one copy of genes on the Z chromosome. As mentioned above, dosage compensation is incomplete in males (the homogametic sex) with two copies of these genes (Garcia-Morales et al., 2015; Zimmer et al., 2016). In **Table S1**, the Z chromosome-specific genes are considered in a separate worksheet. Around 500 transcripts from Z chromosome genes were detected above the expression threshold in macrophages. As reported based upon a smaller set of samples from late-stage embryos, (Zimmer et al., 2016) the large majority of genes on the Z chromosome were expressed more highly in male than in female spleen, although the median ratio was around 1.5-fold rather than twofold suggesting incomplete compensation. Numerous known immune-associated genes (e.g. *CCL19, IL7R, JAK2, CD274, TNFAIP8*), transcription factor genes (*MEF2C, NFIL3*) and metabolic

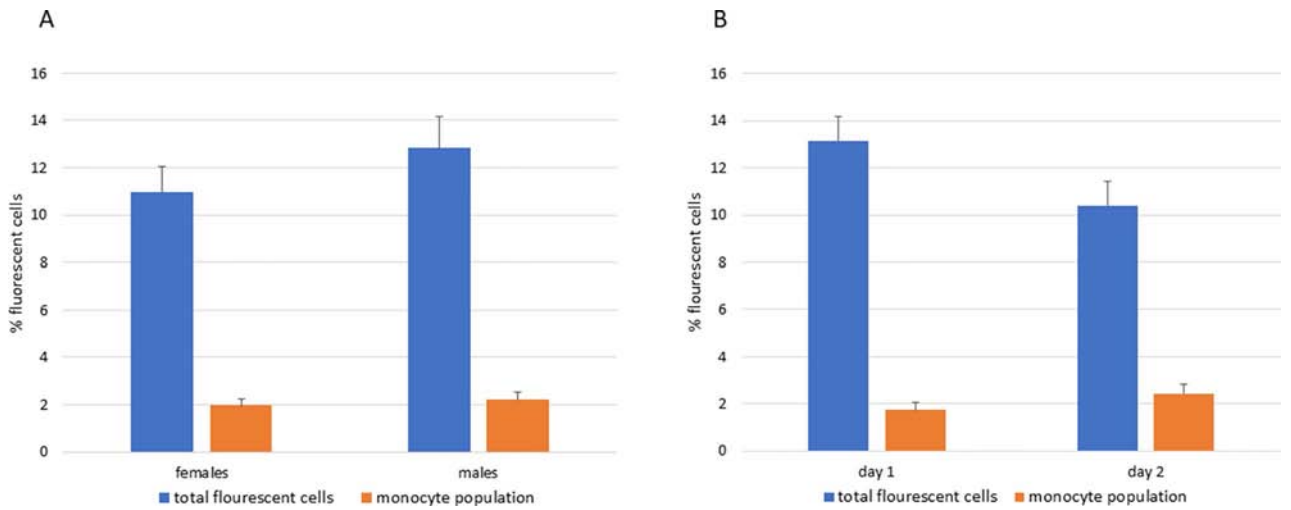


FIGURE 3 | Variation in blood leukocyte populations in F2 progeny of F1 sibling mating. Blood from the entire cohort of hatchling birds was analyzed for total *CSF1R*-mApple reporter gene expression (which measures the myeloid compartment including heterophils) or for blood monocytes (KUL01/mApple+). For logistical reasons these analyses were performed on Day 1 or Day 2 of hatch. Bars show the mean +/- SD. The Figure demonstrates that the variance between birds was small, there was no effect of the sex or day of sampling. **(A)** Blood leukocytes in males and females. **(B)** Blood leukocytes on Day 1 and Day 2 after hatch.

enzyme genes (*ACO1, HMGCR*) on the Z chromosome that were expressed at high levels in spleen appeared to be largely dosage compensated (expression ratio not significantly different from 1). We considered also the possibility that immune-related genes on the Z chromosome might be especially subject to evolutionary selection by virtue of being haploid in females but there was no evidence of variant alleles on the Z chromosome producing more than 2-fold differences in gene expression between F2 individuals. Almost all transcripts varied between individuals across a 2-3 fold range without obvious outliers. The single exception is the antimicrobial gene, avidin, where the level ranged from 5 to 121 TPM.

We next considered the autosomal gene set for spleen. As noted above, these families were created in part based upon the genotype of variants at the *IL34, IL12B* and *IL10* loci. Of these genes, *IL12B* and *IL10* were not expressed above the detection threshold in spleen. *IL34* was expressed, but the level of expression was not correlated with homozygosity for either of the parental SNV alleles. To examine the variability in gene expression, we considered the range of values for individual birds and presented that range as a ratio of maximum to minimum (max/min, giving a fold difference value) (**Table S1**). The max/min ratios for commonly-used reference genes, *ACTB, HPRT* and *GAPDH,* were each around 1.35. The median ratio for the entire data set was 1.6. Of the transcripts for which the max/min ratio was >5, most were known W chromosome-specific transcripts (and others were correlated and likely W chromosome-associated, see below), and the large majority of the remainder had no informative annotation or were annotated specifically as endogenous retroviruses. Recent studies have indicated considerable divergence in endogenous retrovirus insertions in individual birds and provide evidence for their expression in spleen and induction in response to infection (Lee et al., 2017; Qiu et al., 2018; Pettersson and Jern 2019). It is unclear whether the extreme individual variation in expression of these putative retroviral transcripts is functionally important.

To explore whether any of the sets of apparently divergent transcripts were co-regulated, we used the network analysis tool Graphia to generate a sample-to-sample correlation matrix based on expression across the 17 spleen samples, correlated at $r \geq 0.97$. This tool has been used previously in the generation of a chicken transcriptional atlas (Bush et al., 2018). There was no association between the samples based on sex of the bird or genotype for any of the genes of interest. We then generated a gene-to-gene correlation matrix. The resulting network graph at a correlation threshold of 0.8 contained 10,273 nodes (genes) connected by 171,821 edges (connections between nodes of $r \geq 0.8$). The network was clustered using the MCL algorithm with an inflation value of 1.7. The resulting network graph and the annotated clusters are summarized in **Table S2**. In principle, such an analysis might reveal *trans*-acting variation. For example, over-expression of a growth factor or a transcription factor and its downstream targets would be correlated with each other and cluster together. In the case of the three cytokines of interest, there were no clusters that correlated with homozygosity for any of the allelic variants derived from the grandparents. *IL34* was expressed and varied over a 2.3-fold range but was not correlated with any other transcripts including the receptor gene *CSF1R*.

The clustering revealed a subset of clusters that could be ascribed a biological function, which provides an internal control indicating the power of the approach. Cluster 2 contains the large majority of cell-cycle related genes (including key transcription factors *E2F* and *FOXM1*) identified previously in the chicken transcriptional atlas (Bush et al., 2018). As might be expected, they are highly-expressed in hatchling spleen; the narrow range of average of gene expression values across samples reflects relatively small differences in proliferative cell numbers between the animals. Cluster 4 is made up almost entirely of Z

chromosome-associated transcripts, with the average expression around 1.6-fold higher in males than females as discussed above. The small number of transcripts within this cluster that are not Z chromosome-associated are mainly poorly-annotated; only 11 have a current non-Z chromosome assignment. The tight restriction of this cluster to Z chromosome-associated genes indicates that the lack of dosage compensation of genes on the Z chromosome has no downstream impact on expression of transcripts on the autosomes. The reciprocal cluster, Cluster 11, contains W chromosome-specific transcripts that are female-specific in their expression. This cluster also contains multiple candidate protein-coding transcripts that have not been assigned, and only 7 that are currently assigned to an autosome. Cluster 6 is clearly enriched for known markers of lymphocytes (e.g. *CD3E, CD4, CD8A*) and lymphocyte-associated transcription factors (*LEF1, STAT4, TCF7*) and presumably reflects subtle variation in relative lymphocyte content of the spleens. Clusters 10 and 13 contain known epithelial (e.g. *KRT14*) and liver (e.g. *ALB*) associated transcripts respectively, probably reflecting minor contaminants in tissue harvesting. Surprisingly, although macrophages are clearly a major component of the cell populations of the spleen, there was no obvious cluster of macrophage-expressed genes. Macrophage marker genes such as *CSF1R* varied only across a 1.6-fold range but did not cluster with each other. This finding is consistent with the analysis of monocyte numbers above indicating that there is very little inter-individual variation.

Cluster 22 was the only cluster that was both highly-expressed and extremely-variable amongst the samples. It contains several heterophil-specific transcripts encoding granule proteins (*CATHL1, CATHL2, LYSG, MIM1, S100A9*), the cytoplasmic protease inhibitor *SERPINB10* (Rychlik et al., 2014) and the IL8 receptor, *CXCR1*. *CATHL2* is very highly-expressed and anti-CATHL2 antibody has been used a marker for heterophils in hatchling spleen (Cuperus et al., 2016). We conclude that the coordinated variation in this cluster reflects profound variation in heterophil number amongst the different samples. There was no evidence within the cluster of a candidate regulator expressed in spleen that might explain the variation. No candidate regulator was evident even if the correlation threshold was lowered. Genes encoding the CXCR1 ligand (*IL8*) and the G-CSF receptor (*CSF3R*) did vary around 5-fold between samples but were not correlated with each other or with their binding partners. As noted below, we did observe a massive variation in regulated expression of the growth factor gene *CSF3* in macrophages which would provide a clear mechanistic explanation for variable heterophil numbers in the spleen. There is published evidence that the heterophil:lymphocyte ratio varies between birds and is highly heritable (Campo and Davila, 2002). It also varies greatly amongst chicken breeds (Bilkova et al., 2017) and avian species (Minias, 2019). We did not observe any corresponding changes in circulating myeloid cells in these birds (**Figure 3**) so this phenomenon appears specific to the spleen.

Several transcripts varied substantially and idiosyncratically but did not form part of larger co-expressed clusters. Of particular interest are the Class II MHC genes *BLB1* and *BLB2*, which are polymorphic in birds and strongly-linked to disease resistance (Parker and Kaufman, 2017). Both transcripts were highly-expressed and varied more than fivefold between individuals. Despite their chromosomal proximity, their expression levels were not correlated with each other. *BLB1* did not form part of a co-expression cluster. It is normally expressed in intestine but variable expression in spleen has been reportedly associated with particular MHC haplotypes (Parker and Kaufman, 2017). *BLB2* clustered with only two other transcripts (Cluster 690): *CD1B*, one of two CD1 genes in the chicken MHC (Salomonsen et al., 2005) and LOC101747454. The latter is also described as *BLB2* in the NCBI database (https://www.ncbi.nlm.nih.gov/gene), but apparently expressed more highly than the gene annotated as *BLB2*. There are additional genes closely-related to *BLB2* around 400kb distal on chromosome 16 (Parker and Kaufman, 2017) that might produce some ambiguity in mapping. By contrast to the class II MHC transcripts, other MHC-associated transcripts (the BLB2-like *DMB2* transcript which is commonly co-expressed with *BLB2* (Parker and Kaufman, 2017), the class 1 MHC transcript *BF1* and the antigen processing transporter gene *TAP2*) varied less than 2-fold between individuals.

Candidate Spleen Null Expression Variants

As noted above, most transcripts with extreme ranges of expression lacked informative annotation and may be expressed retroviruses or other non-coding RNA elements. There were few protein-coding transcripts that were absent or minimal in only a subset of birds. The relative absence of such variation supports the conclusion that stop-gain mutations in the founder birds are either false-positives or are not sufficiently severe to drive nonsense-mediated mRNA decay. Individual profiles of transcripts affected by candidate null expression alleles are shown in **Figure S1** and discussed below.

The gene encoding the classical Th2 T cell lymphokine, IL4, was heterogeneously expressed and the profile of variation suggested the existence of a null allele. In three birds, *IL4* mRNA was barely detected, and expression in the remaining birds fell into two groups consistent with 1x and 2x functional alleles. There is little functional data on IL4 in birds, although a recent study reported an anti-IL4 antibody and demonstrated that IL4 can drive alternative functional states in chicken macrophages (Chaudhari et al., 2018). Genetic variation in the region of the *IL4* gene has been associated with feather pecking behavior in layers (Biscarini et al., 2010)

The small MAF-related transcription factor gene, *MAFF*, also shows a spread of expression that is suggestive of Mendelian segregation of a null expression allele. Multiple members of the Maf transcription factor family are expressed during chick embryogenesis with partly over-lapping distributions (Lecoin et al., 2004). In the chicken expression atlas (Bush et al., 2018) *MAFF* is most highly expressed in macrophages. As discussed below, similarly diverse expression was detected in bone marrow-derived macrophage data.

One other example of extreme variation in splenic expression with regulatory potential is *GNAS*, encoding a stimulating subunit of G protein coupled receptors, which is a complex imprinted locus in humans. Paternally-inherited mutations in this gene in humans lead to a progressive heterotopic ossification (Bastepe,

2018). Enforced expression of a dominant negative form of GNAS in chicken somites led to rapid ectopic bone and cartilage formation (Cairns et al., 2013). Genetic variation in the *GNAS* region is associated with body weight, muscle meat quality and bone strength QTL in broilers (see Animal QTLdb). The impact of *GNAS* mutation in humans is suggestive of possible roles of GNAS in so-called wooden-breast, a pathology prevalent in high breast-yield broilers (Chen et al., 2019). One other transcript that is highly variable and might conceivably be associated with selection for a production phenotype encodes Islet2 (*ISL2*) which in mice regulates the generation and migration of specific motor neurons (Thaler et al., 2004).

Finally, phosphoserine phosphatase (PSPH) which varied >50 fold between individuals encodes a well-known mediator of L-serine biosynthesis in a variety of tissues. In laying birds, *PSPH*mRNA and protein levels were reportedly increased in the glandular and luminal epithelial cells in the developing oviduct of chicks treated with exogenous estrogen (Lee et al., 2015).

Analysis of Variation in Gene Expression in Bone Marrow-Derived Macrophages Derived From Adult Birds of Parental Broiler and Layer Lines

The advantage of using BMDM for the current purpose is three-fold. Firstly, by contrast to spleen, they are a relatively pure cell population, which increases the sensitivity of detection and should reduce the likelihood of detecting changes in cell populations as opposed to allelic regulation of gene expression in RNAseq data. Secondly, the prolonged *in vitro* culture under defined conditions reduces the potential impact of environmental variation so that differences are more likely due to genetic variation. And finally, these cells respond to stimulation with the TLR4 agonist, lipopolysaccharide (LPS) with a profound change in gene expression. We can therefore monitor the impact of genotype on inducible genes. In a previous study, we were able to identify the differential expression of Z chromosome-associated transcripts in male versus female BMDM, and we inferred that macrophages from female birds have a novel mechanism to compensate for the presence of the inducible interferon genes on the Z chromosome (Garcia-Morales et al., 2015).

The previous study (Garcia-Morales et al., 2015) used gene expression microarrays to quantify mRNA levels and was limited by the available annotation at the time. We first sought to repeat the earlier study and to compare males and females and commercial layers and broilers. The analysis of outbred commercial birds from the parental lines provides a control for the greater expression diversity that we anticipate in F2 birds from deliberate inbreeding. BMDM from three adult female and three adult male broilers and three adult layer females were cultivated with or without LPS for 24h. The primary data are provided in **Table S3**. The prolonged incubation was chosen to avoid temporal differences in the rate of response and specifically to focus on late-response genes that require stimulation by autocrine interferon signaling. One disadvantage is that the dataset does not capture the acute pro-inflammatory and anti-inflammatory transcripts that are induced transiently by LPS. This transiently-induced set includes the negative feedback regulator IL10. We showed previously that autocrine IL10 inhibits LPS-inducible cytokine production in BMDM (Wu et al., 2016). *IL10* was only detected at low levels in the 24 hour-stimulated BMDM (<10 TPM) where the receptor gene, *IL10RA,* was highly-expressed and further induced by LPS.

The expression of known macrophage-specific genes (Bush et al., 2018) including *CSF1R* and the transcription factor *SPI1* was high and invariant among all the samples, supporting the consistency and relative purity of these macrophage populations. Amongst the averaged data for the 8461 transcripts that were expressed >10 TPM in at least 1 BMDM sample, 872 were induced >2-fold and 697 were repressed >2-fold by LPS. The most highly-inducible genes included LPS-responsive transcription factor genes *BATF3*, *HIF1A, IRF1, IRF8* and feedback regulators including *CISH, SOCS3* and *TNFAIP3.* Chicken BMDM, like BMDM from rodents, take up arginine and produce large amounts of nitric oxide in response to LPS (Wu et al., 2016). Accordingly, *NOS2* and *GCH1* (which is required to generate the NOS2 co-factor tetrahydrobiopterin) and the citrulline-arginine recycling enzymes, *ASL2* and *ASS1,* were each induced by LPS in all cultures. Unlike rodent BMDM, which induce the cationic arginine transporter *SLC7A2* in response to LPS (Young et al., 2018) chicken BMDM induced a distinct arginine transporter, *SLC7A3.* One other important difference between chicken and rodent BMDM is the expression of *IRG1,* now annotated as aconitate decarboxylase (*ACOD1*). In mammalian macrophages, *ACOD1* was profoundly-induced by LPS. Inducible *ACOD1* has been attributed roles in metabolic reprogramming in stimulated macrophages, subverting the TCA cycle by diverting iso-citrate and catalyzing the generation of cis-itaconate, which has a proposed anti-inflammatory feedback function (Mills et al., 2018). In the chicken BMDM, *ACOD1* was already highly-expressed in the unstimulated state albeit induced further by LPS. Interestingly, *ACOD1* polymorphism has been linked to resistance to the macrophage-tropic pathogen Marek's disease virus (Smith et al., 2011). One transcript of particular interest is *IGF1*, encoding a major regulator of somatic growth which is associated with a signature of selection in broilers (Qanbari et al., 2019). In mammals, *IGF1* is highly-expressed in macrophages and regulated by CSF1 (Gow et al., 2010) but the expression in chicken BMDM was below the detection limit.

The LPS-repressed genes include cell cycle-associated transcripts, such as *BUB1, FOXM1* and *MKI67* (which encodes the commonly-used proliferation marker KI67), reflecting the known ability of LPS to inhibit cell proliferation in this culture system. Associated with this growth inhibition, as in mammalian macrophages (Yue et al., 1993), LPS stimulation down-regulated expression of *CSF1R.* LPS treatment repressed other transcripts encoding multiple cell surface receptors and secreted effectors to a much greater extent than *CSF1R.* High-expression transcripts reduced >10-fold included membrane receptors *TREM2* (and related *TREMB1* and *TREMB2*), *GPR34, ITGB5, MARCO, TLR2A, TLR2B, TLR7,* and *ENSGALG00000028304* (encoding the macrophage mannose receptor/KULO1 antigen (Hu et al., 2019).

Differences in regulation between males and females are considered further below, and in these commercial birds, we

sampled only three male broilers. The expression of transcripts on the Z chromosome is shown on a separate sheet in **Table S3**. It is nevertheless striking that of 380 Z chromosome-encoded transcripts detected in BMDM with expression levels of >10 TPM, <10% were up or down-regulated by LPS.

Table S3 also shows the range of expression for the control and LPS-stimulated samples. There was considerably greater variation than observed in the spleen RNAseq data. 615 transcripts varied across >5-fold range in the unstimulated data and 1110 by >5-fold across the LPS-stimulated samples. To explore this variation further and seek evidence of co-regulated genes we again used Graphia. **Figure 4** shows a sample-to-sample network graph ($r \geq 0.97$) for these data. This shows that there was no segregation of the broilers from layers, suggesting that there is no consistent strain-specific gene expression pattern. Consistent with the evidence of profound LPS-induced changes in gene expression, the main axis of separation is driven by LPS stimulation which was analyzed in more detail in the F2 progeny.

Table S3 also summarizes the fold changes comparing the broilers and layers in the control and LPS-stimulated states. Consistent with the network analysis, there are few annotated transcripts, and even fewer highly-expressed transcripts, that distinguish the expression profiles based upon breed. The most broiler-enriched transcripts of interest are *S100A8, CCL5* and *ETS2*, whilst the layers had higher expression of *CMPK2, SLC40A1, APOA1, IFIT5, STAT1* and *MMP9.*

Analysis of Gene Expression in BMDM Generated From Progeny of F1 Brother-Sister Matings

To continue to address the hypothesis that sibling-mating would expose homozygosity for high and low expression alleles and amplify the variation seen in the commercial broiler and layer lines, we isolated bone marrow from a total of 32 hatchlings from different families, grew BMDM, treated them with or without LPS as above for 24h, isolated mRNA and profiled gene expression by RNAseq. As with spleen, we chose hatchlings to avoid possible confounding influences of pathogen exposure including routine immunization with live vaccines. We also hoped to validate a method that would enable early and rapid screening of the progeny of defined matings that might form the basis of breeding decisions. A sample-to-sample analysis of the complete dataset was performed using Graphia and four samples from stimulated and unstimulated states were identified as major outliers and were removed. To enable pairwise-comparisons of stimulated and unstimulated states, we further removed those samples for which there was not a pair, leaving a total of 28 F2 birds from 6 separate families for analysis (+/- LPS). This is a proof-of-concept experiment and it was beyond our resources to survey the full genetic diversity within the founders. To maximize the likelihood of detecting multiple F2 birds with the same expression variant inherited from a grandparent mating we included 16 birds from 4 F1 brother-sister matings from the same grandparent cross. They are in effect double-cousins. We also included six birds from family H (which exhibited poor hatch rate, low weight at hatch and poor survival (**Figure 2**) to explore

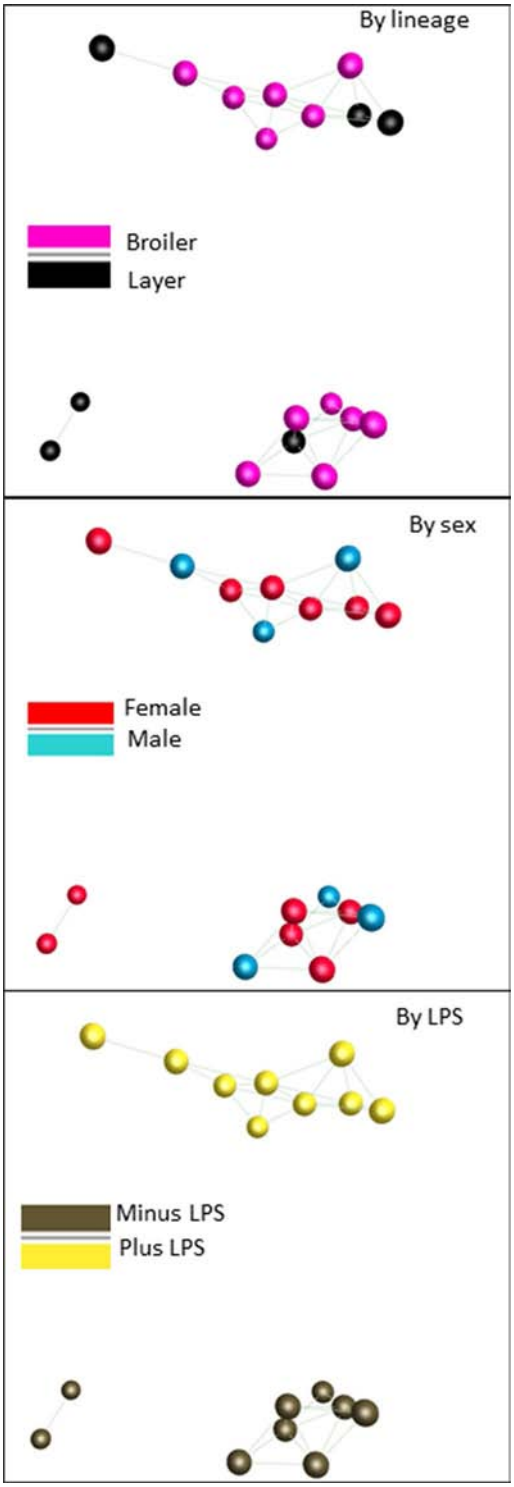


FIGURE 4 | Network analysis of the response of BMDM generated from adult broiler and layer chickens to LPS. RNAseq gene expression data from BMDM cultured with or without LPS for 24h was analyzed using the network visualization tool Graphia. The panels show the sample-to-sample matrix with a Pearson correlation threshold of 0.97. In the three panels birds are identified based upon breed, sex or treatment as indicated. Note that the samples segregate solely based upon LPS stimulation. The sets of LPS-regulated transcripts that contribute to this separation are summarized in **Table S3**.

possible detection of deleterious variants in this sibship, and a smaller number of birds from other grandparental matings to include broiler-layer specific variant *CSF1R* alleles.

Initial analysis of the expression data revealed variable detection of multiple genes associated with mesenchymal lineages including numerous collagen genes. These transcripts were not detected in the BMDM from adult commercial birds in **Table S3**. Neonatal calvarial cultures are routinely used to generate osteoblasts in mice. Such cultures contain large numbers of macrophages even without addition of growth factors (Chang et al., 2008). The FANTOM consortium recently published an analysis of transcriptional regulation of chicken promoter during development, including a sample annotated as bone marrow-derived mesenchymal stem cells which is actually a hatchling calvarial bone marrow culture (Lizio et al., 2017). This sample exhibited abundant expression of known macrophage-specific genes including *CSF1R* alongside multiple collagens. Accordingly, we concluded that the hatchling bone marrow (unlike adult) probably contains mesenchymal stem cells which proliferated and differentiated alongside the macrophages in our culture system. In mouse calvarial cultures the macrophages and osteoblasts interact with each other to control calcification, an interaction that is paralleled *in vivo* (Chang et al., 2008). The culture system therefore unexpectedly enabled us to examine possible mesenchyme-associated gene expression variants that are quite likely relevant to the phenotypic diversity in the broiler-layer cross, but such an analysis required deconvolution of the data.

We first considered the sets of control and LPS-stimulated samples separately (**Table S4**) and calculated the ratio of maximum/minimum expression. The extent of variation amongst protein-coding transcripts was massively greater than in the spleen or adult broiler and layer BMDM data in **Table S3**. To identify possible null (absolute loss of expression) variation, we identified the set of transcripts for which the maximum was >20 and minimum <1. **Table S4** includes a Venn diagram for the control and LPS-stimulated states. Of a total of 962 transcripts that met the criterion for extreme variation between individuals, 365 (39%) overlapped between the two sets (control, + LPS) and 432 (45%) were specific for the LPS-stimulated state. The set of variable expression transcripts that is independent of LPS stimulation is clearly enriched for mesenchyme-associated genes including 10 separate collagen genes. Some of these transcripts (e.g. *COL1A1*, *COL1A2*) appear in the LPS-stimulated list only because they are marginally above the detection limit (>1 TPM) in the lowest-expressing sample. Nevertheless, the analysis validates some of the conclusions from the spleen data suggesting the existence of effective null expression alleles for GNAS, *IL4* and *MAFF*. Furthermore, *CSF3,* which was profoundly LPS-inducible in the large majority of birds, was barely detected in others. Such variation could contribute to extreme variation in heterophils in the spleen (**Table S2**).

The expression of *CSF1R* mRNA in the F2 hatchling birds was around 30% lower than in the BMDM cultures from adult commercial birds (221 versus 321 average TPM), but down-regulated to a similar value (145 versus 132 average TPM) in the LPS stimulated cultures. The level of *CSF1R* mRNA varied over a much greater range (47-347 TPM) in the F2 BMDM cultures compared to the commercial birds.

We next deconvoluted the data by network analysis using Graphia. We anticipated that transcripts associated specifically with gene expression in the separate mesenchyme and macrophage populations would form separate clusters and their relative levels would provide a surrogate for the relative purity of each cell culture. Network graphs for this dataset are shown in **Figure 5**. The sample-to-sample profile (**Figure 5A**) is color-coded for family (left), sex (middle) or treatment (right). The samples did not separate based upon sex and there was also no obvious segregation based upon family or the parent allelic variants selected for analysis. By contrast to the BMDM data from the commercial birds there was also no separation based on LPS stimulation. The average profiles of the largest clusters derived from the gene-to-gene analysis ($r \geq 0.85$) are shown in **Figure 5B**. Key genes and functional annotation terms for the larger clusters are summarized in **Table 3** and the full gene lists in each cluster are provided in **Table S5**.

Mesenchyme-Related Gene Expression in F2 Bone Marrow Cultures

Cluster 1 which contains 1396 genes, includes major bone-associated collagens (*COL1A1, COL1A2*) and extracellular matrix proteins alongside multiple cell cycle-associated transcripts, and is enriched for GO terms associated with mesenchyme and extracellular matrix. This cluster of genes most likely reflects the presence of varying numbers of proliferating mesenchymal cells. There are around 180 transcripts with no current informative annotation which can be inferred to be related to mesenchyme differentiation. The expression of mesenchyme-associated transcripts was not regulated by addition of LPS in any of the birds. Given that one grandparent was a broiler, we suggest that differential growth of mesenchymal cells in culture is related to selection for growth and muscle/bone/fat related production traits. If there is a genetic basis for the variable growth of mesenchymal cells in this culture system, it is likely to involve regulated expression of growth factors or growth factor responsiveness. Transcripts encoding members of each of the many families of growth factors implicated in mesenchymal stem cell (MSC) growth and differentiation (e.g. *BMP4, BMP6, CTGF, FGF13, INHBA, NOTCH2, PDGFB, TGFB3, VEGFA,* and *WNT5B*) were each highly-expressed and varied greatly between individual birds (**Table S5**). Most were not contained within cluster 1, but the cluster does contain transcripts encoding multiple growth factor receptors (e.g. *ACVR1, ACVR2A, DDR2, EPHA3, EGFR, PDGFRB, SMO, TGFBR2)* and an equally plausible mechanism is regulated expression of these receptors. Cluster 1 also contains *TGFB3* and *THBS1* (thrombospondin 1) both of which control MSC proliferation in humans (Belotti et al., 2016). Genes within this cluster are candidates for causal association with broiler production traits. Consistent with that view, *ASPH* lies within a QTL interval associated with muscle development on chromosome 2 (Godoy et al., 2015) and *THBS2* has been identified as a candidate gene within a QTL for fatness in chicken (Moreira et al., 2015).

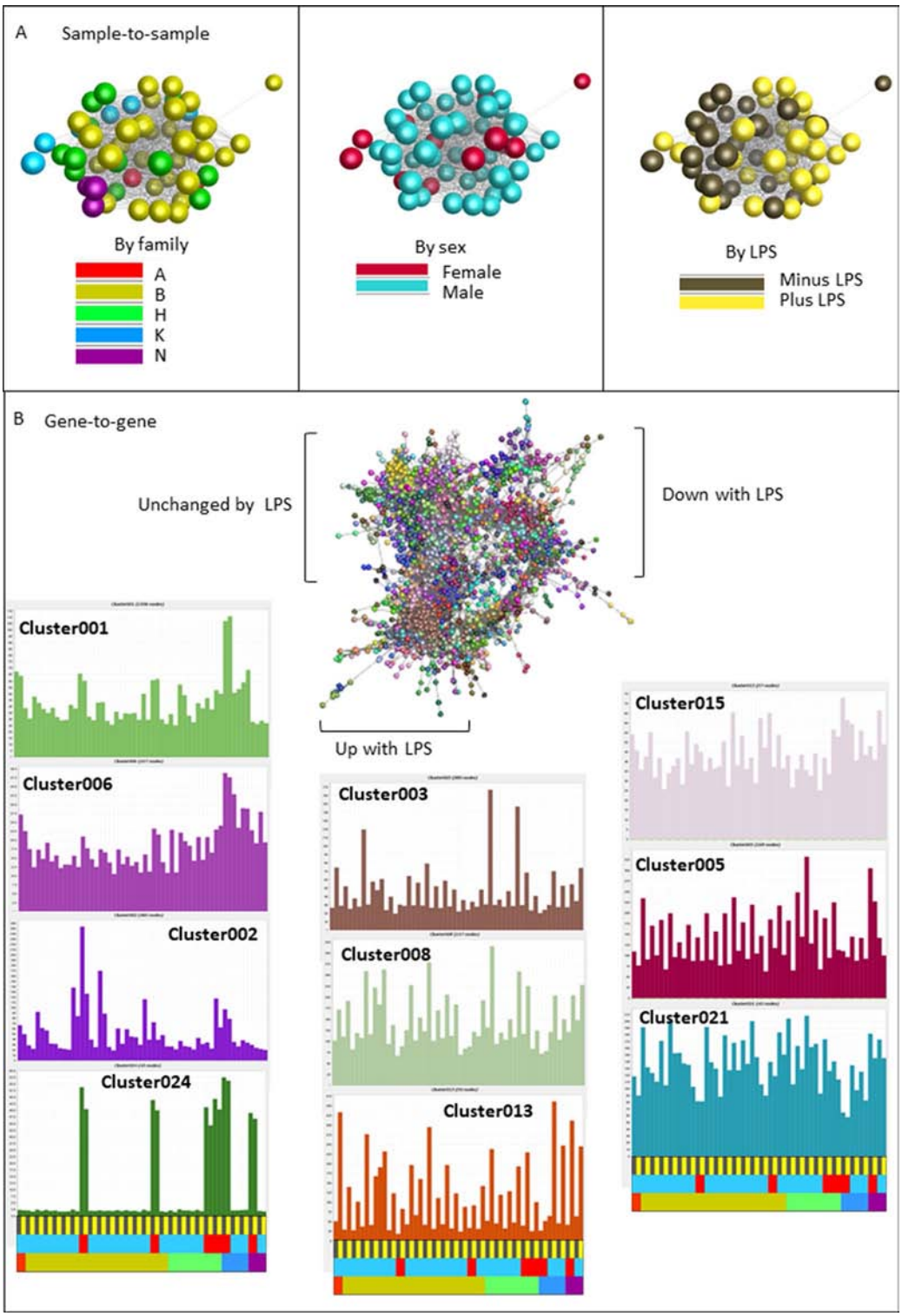


FIGURE 5 | Network analysis of the response of BMDM generated from F2 inbred hatchling chickens to LPS. RNAseq gene expression data from BMDM cultured with or without LPS for 24h was analyzed using the network visualization tool Graphia. (Panel **A**) shows the sample-to-sample matrix at a Pearson correlation threshold of 0.97. Note that there is no clear separation based upon family or sex, and less segregation based upon LPS treatment than in the parental comparison (**Figure 4**). (Panel **B**) shows the gene-to-gene matrix generated at Pearson $r \geq 0.85$ with clusters of co-regulated transcripts colored. This analysis reveals a clear segregation of clusters that are increased, decreased or unchanged by LPS. The average profiles of selected clusters discussed in the text are shown in the surrounding histograms. The color code at the bottom of each column indicates LPS versus control in pairs from the same birds, sex, or family (colors as in Panel **A**). The transcripts contained within each Cluster are shown in **Table S4** and **Table S5**.

TABLE 3 | Functional annotation of clusters derived from untreated and LPS treated F2 BMDM.

Cluster Number (number of nodes)	Expression pattern	Representative genes	Enhanced terms (P value)
001 (1396)	Unchanged by LPS	*ACTA2, ADAMTSs, ASPH, COL1A1, COL1A2, FBN1, FN1, MFAPs, TGFB3, TGFBR2/3, THBS1, THBS2*	Protein processing in endoplasmic reticulum (5.5E-10); Poly(A) RNA binding (3.2E-8); Extracellular matrix (1.6E-6); Endoplasmic reticulum (8.7E-6); ER-Golgi transport (7.5E-5); Focal adhesion (1.5E-4)
002 (485)	Unchanged by LPS	*COL9A1, COL10A1, COL11A1, FGFRs, FOXO6, IRX5/6, MEF2A/B/C, PAX3, RUNX2, SIX2/3, SMAD1, SOX5/8/9*	Extracellular matrix (0.003); Secreted (0.007); Glycoprotein (0.012); Signal (0.038)
003 (389)	Up with LPS	*ATF3, BHLHE40, CD274, CEBPB, FOSL2, GGCL1, GVIN1, IFIT5, IFIH1, IRF1, IRF8, IRF9, OASL, SOCS1*	Zinc finger, RING/FYVE/PHD-type (0.02); Influenza A (0.01)
004 (259)	Unchanged with LPS	*CPSF2, DDX6, LEO1, PAPD5, PAPOLA, PARN, PRPF38B, PRPF39, PRPF40A, PRPF6, SMU1, TOP2B, TOP3B, TOPBP1*	Spliceosome (2.4E-9); Cytoplasmic mRNA processing body (0.0004)
005 (168)	Down with LPS	*C3AR1, CD14, FAAH, MAOA, PEX16, PSEN2, TNFAIP8L1, TUBB1*	Transmembrane helix (0.1); Aldolase-type TIM barrel (0.07)
006 (167)	Unchanged by LPS	*BUB1, BUB1B, BUB3,* centromere protein genes, kinesin genes, kinetochore complex components	Cell cycle (4.9E-14); Mitosis (9.8E-13); Cell division (2.3E-11); Centromere (8.3E-11)
007 (160)	Unchanged by LPS	*RPL* and *RPS* genes, translation initiation and elongation genes	Structural constituent of ribosome (3.6E-81); Ribosome (4.7E-79); Translation (1.2E-68); Protein biosynthesis (1.8E-7)
008 (137)	Up with LPS	*BACH1, IRF6, KLF8, MAFK, REL, TLR15*	No significant enrichment
009 (100)	Down with LPS	*IGF2, KLF3, KLF13, MAF1, TAF12*	No significant enhancement
010 (76)	Unchanged by LPS	*AAR2, CNOT10, EXOC1, GABPA, POLR1A*	RNA recognition motif (0.002)
011 (71)	Unchanged by LPS but one animal very high	*ACTG2, CDH11, COL7A1, ENPP2/3, FGF7, NPY, PDGFRA, VAV2*	Calcium ion binding (0.2)
012 (65)	No trend with LPS	All unannotated *ENSGALG*	
013 (59)	Up with LPS	*CCL4, BATF6, FLT1, IL13RA1, IL1B, NFKB1, NFKB2, NFKBIA, TRAF2, TRAF3, ZC3H12A*	Toll-like receptor signaling pathway (0.0005); Cytosolic DNA sensing pathway (0.001); RIG-I-like receptor signaling pathway (0.001); Inflammatory response (0.01)
014 (57)	Up with LPS	*CCNE1, FMR1, SMCHD1, TRAF6, USP16*	Cell cycle (0.1)
015 (57)	Down with LPS	Mostly unannotated *ENSGALG* and *LOC* genes	No significant enhancement
016 (52)	Down with LPS	*ATP6AP1, ATP6AP2, ATP6V0D1, CASP9*	No significant enrichment
017 (52)	Unchanged with LPS	*ACTR1A, COG5, E4F1, EIF2AK4, INSIG2*	No significant enrichment
018 (52)	Variable response to LPS	All unannotated *ENSGALG*	
019 (51)	Variable response to LPS	*ACOD1* (*IRG1*), *IRF2, MAP2K3, TAOK1, TBK1, USP15*	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process (0.01)
020 (45)	Unchanged with LPS	*CCT* genes, *HSP* genes	Positive regulation of protein localization to Cajal body (1.5E-10); Chaperonin TCP-1, conserved site (4.7E-10); Positive regulation of establishment of protein localization to telomere (2.7E-10); Chaperone (2.1E-9)
021 (45)	Down with LPS in most	*CSF1R, IL2RG, IL6R, TLR4*	No significant enrichment

Benjamini Hochberg corrected P values are presented. First 21 clusters only as number of genes becomes too low for meaningful analysis in smaller clusters. This analysis used DAVID (https://david.ncifcrf.gov/home.jsp) to determine enrichment for annotation terms.

Cluster 2 is a distinct mesenchyme cluster that varies to a much greater extent between birds than cluster 1, and in the majority of birds the average expression of transcripts within the cluster was down-regulated by LPS. The GO enrichment indicates an association with extracellular matrix and secreted proteins, including collagens associated specifically with hypertrophic chondrocytes (e.g. *COL9A1*, *COL10A1* and *COL11A1*), four FGF receptor family members, and many other transcripts associated with hypertrophic chondrocytes in mammals. Cluster 2 also contains many known transcriptional regulators of chondrocyte development in mammals [reviewed in (Liu et al., 2017)] including *IRX5* and *IRX6, MEF2A, B* and *C*, *PAX3, RUNX2, SIX2* and *SIX3*, *SMAD1, SOX5, SOX8* and *SOX9*. The coordinated regulation of these factors suggests that the basic biology of chondrocyte differentiation is conserved in birds, but also that the individual birds/culture differ greatly in their support of this pathway. There is at least one obvious candidate regulator in this cluster. Chondromodulin (*CNMD*) is required for the maturation of chondrocytes and expression of *Col10a1* in mice (Yukata et al., 2008) and was amongst the most divergent transcripts in the F2 cultures (**Table S5**). We suggest that genes within clusters 1 and 2 are likely candidates underlying growth and composition traits in broilers.

Variable Expression of LPS-Regulated Genes in BMDM From F2 Inbred Birds

LPS signals to macrophages through the receptor TLR4. The response of macrophages to LPS involves two distinct adaptor proteins, MYD88/TIRAP and TRIF/TRAM and downstream target genes can be classified based upon their dependence on these two effector pathways. The TRIF/TRAM pathway links to induction of type 1 interferon (IFN) and an autocrine stimulatory cascade. Until recently it was claimed that LPS-stimulated chicken macrophages do not produce endogenous

IFN. However, a recent study (Ahmed-Hassan et al., 2018) reported the release of type 1 IFN activity from LPS-stimulated macrophages and provided evidence of autocrine signaling. They were not able to detect the induction of *IFNB* mRNA. *TLR4* is highly polymorphic at the protein-coding level in chicken (Swiderska et al., 2018). Our F2 dataset contains one bird (B628) in which *TLR4* mRNA was exceptionally low in both control and LPS-stimulated states. *TLR4* is part of a small cluster (cluster 21) that also includes *CSF1R* but does not include any other macrophage-specific genes. Neither *CD14* (encoding the TLR4 co-receptor), nor *MYD88* or *TIRAP* varied substantially between birds. Accordingly, we conclude that this bird was specifically deficient in *TLR4* expression.

The candidate targets of the TRIF/TRAM pathway are regulated by transcription factors of the IRF family. Cluster 3 contains the largest set of transcripts that was up-regulated by LPS. It is much smaller than the equivalent in commercial birds because of tighter correlations generated by the larger dataset, and because the level of induction varied substantially between individuals. The likely driver of the variation between birds is differential regulation of the LPS-inducible transcription factors, *IRF1, IRF8* and *IRF9* as well as *ATF3, BHLHE40, CEBPB* and *FOSL2*. In human monocytes, eQTL analysis of the LPS-inducible gene expression response revealed *trans*-acting variants impacting upon the regulation of the suite of genes regulated by inducible autocrine IFN signals (Fairfax et al., 2014). *IRF1* and *IRF8* are obligatory intermediates in induction of many LPS-inducible genes in mouse macrophages ((Roy et al., 2015) and references therein) and in chickens *IRF1* is a key mediator of type 1 IFN antiviral signaling (Liu et al., 2018). Cluster 3 contains many known IFN-inducible effector genes amongst which the most highly-expressed/inducible and hypervariable include C*D274* (encoding the check point inhibitor PD-L1*), GVIN1, GGCL1, IFIT5, IFIH1, OASL* and the feedback inhibitor *SOCS1*. We can infer that the other transcripts in this cluster, including those that are poorly annotated, form part of the IFN effector system. *IRF1* mRNA was induced in all but one of the birds, but the stimulated level of expression varied 250-fold amongst birds. The induction of *IRF1* in mouse and human macrophages is triggered by transient expression and autocrine signaling by IFNB1, acting through the type 1 interferon receptors, IFNAR1 and IFNAR2 (Sheikh et al., 2014). Both IFN receptor genes were robustly expressed in all the BMDM preparations with relatively little variation. However, birds lack *IRF3*, the key upstream regulator of *IFNB1* induction in mammals, and some evidence places *IRF1* upstream of *IFNB1* induction in chickens (Liu et al., 2018).

There are several other clusters containing IFN-associated genes that are distinct from cluster 3. Cluster 8 contains *IRF6* alongside other inducible candidate transcriptional regulators, *BACH1, KLF8, MAFK* and *REL*. One notable gene within this cluster is *TLR15*, a member of the TLR1 family that is unique to birds and reptiles and implicated in response to fungal/yeast pathogens (Boyd et al., 2012). Most transcripts in cluster 19 were expressed constitutively and induced further by LPS. This cluster includes *IRF2,* transcripts encoding signaling molecules (*TAOK1, TBK1, MAP2K3*) and the feedback inhibitor *USP15*. Cluster 19 contains the profoundly-inducible *IRG1* gene (now annotated as *ACOD1*) discussed above. The transcriptional regulator *IRF7* is in a much smaller cluster, cluster 35, along with the genes for both IL23A and IL12B, which form the functional dimer of the proinflammatory cytokine IL23, and the chemokine CCL20.

The IFN genes in chickens are on the Z chromosome and remain poorly annotated (Goossens et al., 2013) and in any case, the pulse of endogenous *IFNB1* that occurs in stimulated mammalian macrophages is transient (Liu et al., 2018) and would not have been captured in this time course. Our data highlight that autocrine IFN induction occurred in the response of chicken BMDM to LPS, but we cannot determine definitively whether the >100-fold variation in IFN target gene expression we observed is in part due to variation in IFN induction. One hint can be gained from the fact that the F2 birds are a mixture of males and females. The individual male and female samples are highlighted in **Table S4** and the male/female ratio in gene expression is calculated. *IRF1* and many of the known target genes were each expressed significantly more highly in the LPS-stimulated cultures of the male birds, suggesting that there is a correlation between their expression and variation in IFN production. In mice *IRF1* and *IRF8* are regulated independently and cooperate in induction of subsets of interferon target genes in macrophages (Langlais et al., 2016). The cluster analysis indicated that *IRF2, IRF6*, and *IRF7* vary independently of *IRF1,* each controlling a subset of IFN target genes. We suggest that there are regulatory variants strongly impacting the IFN pathway at multiple levels. As shown in **Figure 1C**, the broiler and layer founder birds can be distinguished based upon SNPs associated with the IRF loci alone, suggesting that selective breeding has also selected variation at each of these key regulators. **Table S6** lists all of the variants detected in the founder birds at each of the IRF loci. They include 15 SNPs within the 1kb promoter region of *IRF1* and 24 within the 1kb promoter region of *IRF7*. There are also several non-synonymous coding variants in *IRF7* that are enriched in the broilers but are not predicted to be deleterious.

MYD88/TIRAP1 activation is connected through IRAK1/4, TRAF6 and the kinase TAK1 to activation of the transcription factor NFKB. Most NFKB target genes in macrophages are induced transiently and subsequently repressed by the combined actions of numerous feedback repressors (Baillie et al., 2017). Cluster 13 is the largest cluster that contains LPS-inducible transcripts likely induced by the MYD88/TIRAP-dependent pathway and includes the pro-inflammatory cytokine gene *IL1B* and the chemokine gene *CCL4*. This cluster contains transcripts encoding a number of known feedback regulators of the response including *BATF3, NFKIA*, and *ZC3H12A*. As evident from the averaged profile, the level of induced expression of this cluster of genes is much less variable between individuals. It is also not significantly different between the males and females. Hence, there is clearly separate regulation of the two gene sets lying downstream of MYD88/TIRAP and TRIF/TRAM.

The reciprocal to the LPS-inducible gene sets is the clusters of genes that are repressed by LPS. These clusters are not highly

variable between individuals. The GO term enrichments are summarized in **Table 3**. Cluster 6 is very significantly enriched for known cell cycle-related transcripts and cluster 7 for components of the ribosome and translation apparatus, both likely reflecting inhibitory effects of LPS on cell proliferation in macrophages, and possibly also in the mesenchyme component of the culture.

Although the macrophage content of the cultures likely varies inversely with the mesenchymal content, we do not detect a macrophage-specific cluster. *CSF1R* forms part of small cluster (cluster 21) alongside *TLR4* that is, as expected, also down-regulated by LPS, whilst the core macrophage-specific transcription factor gene, *SPI1* does not correlate with any target genes (Cluster 80). We interpret this to indicate that there is considerable cis-acting variation in the large majority of macrophage-specific genes so that correlations with other macrophage-expressed transcripts fall below the threshold chosen.

Analysis of the Low Fitness F2 Family (Family H)

As mentioned above, one family showed consistently low weight at hatch (**Figure 2**). This family also had a low hatch rate and three individuals died at or just before hatch with abnormalities in the brain, cartilage and muscle seen at autopsy. We compared gene expression in BMDM from this family to that in all other F2 birds, before and after exposure to LPS, to determine whether this family showed altered expression patterns. In the untreated samples, DAVID analysis showed that genes that were lower in Family H included those with GO terms relating to signal transduction, skeletal system development, collagen, extracellular matrix, chondrocyte differentiation. GATHER analysis confirmed the association with skeletal and cartilage development and cell signaling. Since expression of connective tissue genes reflects in part contamination of BMDM with mesenchymal cells (as discussed in detail above), these results may suggest a deficiency in connective tissue formation or the interaction of these cells with macrophages. For genes that were higher in Family H, DAVID found slight enrichment for GO terms associated with receptors, while GATHER detected GO terms associated with response to stimulus, defense response and immune cell activation.

After LPS stimulation of BMDM, there was enrichment for immunoglobulin and extracellular matrix GO terms and GATHER found an association with cell signaling, metabolism and ion transport in the genes that were lower in Family H than the other samples. Among genes that were higher in Family H after LPS stimulation there was enrichment for terms related to the response to LPS and apoptosis (DAVID analysis). GATHER also found association with terms for cell death/apoptosis as well as terms relating to bone formation. The results are presented in **Table S7**.

These results suggest that the poor hatch rate, low hatch weight and early mortality in Family H was related to a failure of normal development due to low expression of key morphological genes which may indicate a deficiency of mesenchymal cells in the bone marrow. There may have been concomitant infection with perturbed expression of genes of the immune system. The Family H BMDMs may have had a greater response to LPS since GO terms associated with the response were higher in these cells.

Candidate Null Expression Alleles

The set of candidate genes showing extremes of expression (including candidate nulls) in **Table S4** clearly reflects in part the variable contribution of mesenchyme lineage cells to the cultures. However, there are a number of genes that were highly-expressed, or selectively-induced in an entirely gene-specific manner. Selected examples are shown in **Figure S2**, which illustrates that each has a different pattern of variation between individuals. *ENSGAL00000028304* was recently shown to encode a macrophage mannose receptor (MRC1L-B, now annotated as MMR1L4), recognized by antibody KUL01 (Staines et al., 2014). KUL01 is a widely-used monocyte-macrophage marker. There are multiple members of this family encoded by the chicken genome, but *MRC1L4* is the only one expressed in BMDM. It forms part of a small cluster (cluster 48) that also includes *TLR2A*. Pentraxin 3 (PTX3) in mammals is required for effective host defense against influenza infection (Reading et al., 2008). In chickens PTX3 is as an acute phase marker of bacterial infection that was undetectable in spleen (consistent with our data) but induced rapidly by infection (Burkhardt et al., 2019). *PTX3* was highly-expressed in control BMDM and regulated only marginally by LPS. The expression varied over 3 orders of magnitude between individuals and was not correlated with any other gene.

Previous studies of induction of pro-inflammatory cytokines in heterophils of birds selected for resistance to *Salmonella* revealed a positive correlation between resistance and induction of *IL6* and *IL8* (Swaggerty et al., 2004). The induced levels of these two genes also varied greatly between birds in our study and were not correlated with each other. There have, to our knowledge, been no published studies of *CSF3* regulation in chickens. *CSF3* was massively induced by LPS and clustered with putative IRF1/8/9 target genes in cluster 3. The extent of variation, with several birds showing almost undetectable expression, indicates that there are either *cis*-acting variants or that *CSF3* is regulated by multiple transcription factors within cluster 3. Regardless of mechanism, *CSF3* variation likely underlies variation in the heterophil response to infection.

We noted in the discussion of splenic gene expression that the class II MHC gene, *BLB2* showed evidence of variation between individuals. Chicken BMDM are strongly Class II MHC-positive but *BLB2* expression varied between birds. *BLB2* expression varied in parallel with *CD74*, which encodes the class II-associated invariant chain. Since these are on different chromosomes, there is likely to be an upstream *trans*-acting regulator. The obvious candidate, *CIITA*, was not detected in our annotation. Two other genes of interest were

IL4 and *IL34*. In keeping with the findings from the spleen discussed above, *IL4* was detected in a subset of preparations, unaffected by LPS, and undetectable in others. *IL34* was barely detectable in BMDM. *CSF1,* encoding the macrophage growth factor, was expressed constitutively and induced by LPS, but was also very variable amongst samples.

CONCLUSION

The main purpose of this study was to demonstrate that allelic imbalance in gene expression and/or coding variants of large effect in an outbred commercial chicken population could be uncovered by brother-sister mating of F1 progeny to generate an array of F2 individuals, and to survey the impact of homozygosity for the possible set of variants. It was not our intention to document such variants extensively. The founder birds were chosen deliberately to be as outbred as possible and as different from each other as possible.

Despite the high prevalence of candidate null mutations in the grandparental broilers and layers, we found little evidence of adverse impacts of sibling mating on growth and development. It is likely that the highly-selective breeding of commercial birds has largely purged variants that impact on hatch-rate and fertility. Extensive inbreeding in pedigree dogs is not strongly-associated with health-related traits (Jansson and Laikre, 2014). Similarly, a large-scale survey of Irish cattle genotypes revealed relatively few examples of severely deleterious recessive alleles affecting survival or production traits (Jenko et al., 2019). A growing literature in humans, based upon analysis of populations where consanguineous marriage is common, shows that the large majority of the thousands of homozygous null mutations detected in such populations have no overt phenotype (Erzurumluoglu et al., 2016). However, if the gene function is known, more subtle phenotypes can be detected (Saleheen et al., 2017).

Like the direct measurement of gene products and their functions, analysis of mRNA levels provides an intermediate phenotype to assess the impact of homozygosity. The variation we observed in the hatchling spleen supported the hypothesis that there are allelic variants in parental broiler and layer lines that strongly impact on gene expression in immune cells. This also indicated a major difference in heterophil accumulation in the spleen evident from the correlated expression of known marker genes. In the analysis of BMDM, by comparison to the set of parental outbred commercial broiler and layer lines, the F2 birds exhibited considerably greater variation in both basal and LPS-inducible gene expression. In particular, we highlighted a set of interferon-inducible genes that was co-regulated and varied in expression across a 100-1000-fold range. We infer that this variation is associated with the extensive polymorphism in IRF family members (**Figure 1B**) that distinguishes broilers and layers. The approach we have demonstrated has some potential for use in selective breeding since it does not require mature birds or disease challenge. It is entirely plausible that the generation of homozygosity for low expression alleles that we have identified contributes to the phenomenon of inbreeding depression (Charlesworth and Willis, 2009).

If there is, as we suspect, genetic variation that impacts on IFN production encoded by the Z chromosome, the analysis of BMDM from female hatchlings could be applied as progeny testing to select high and low responder lines, assaying only for the LPS-inducible target genes in Cluster 3 from the analysis of BMDM. The culture system could also be used to assess candidate genes and the impact of putative *cis*-acting variation. For example, in the Sal1 locus affecting *Salmonella* resistance in birds, two candidate genes have been identified, *AKT1* and *SIVA1* (Psifidi et al., 2018). *SIVA1* was not detectably expressed in BMDM. Several other genomic regions underlying QTL for *Salmonella* resistance in this and previous studies [e.g. (Thanh-Son et al., 2012)], contain genes that were hypervariable amongst F2 progeny (e.g. *IRF1* and *IRF6*) in the dataset analyzed here. In principle, using the founder DNA and mRNA sequences we could infer the existence of allelic homozygosity for each of the transcripts detected in the spleen and BMDM RNAseq datasets but the size of the population analyzed here was not sufficient to separate *cis*-acting from *trans*-acting variation or to attribute causation. It remains to be seen whether the variation in bone marrow-derived mesenchyme proliferation that we uncovered *inter alia* might also provide an opportunity to accelerate phenotypic selection for production traits.

DATA AVAILABILITY STATEMENT

The variant call files (VCF) for the genomic DNA sequence are available at Edinburgh Datashare (datashare.is.ed.ac.uk) under the Data title: "Variant discovery from whole genome sequence data from a commercial broiler line and a CSFIR-mApple reporter transgenic layer line." The primary RNAseq sequence data have been deposited in the European Nucleotide Archive under Bioprojects PRJEB22373 and PRJEB34093. The primary and processed DNA and RNA sequence data will also be made available by the authors, without constraint on use, to any qualified researcher upon request.

ETHICS STATEMENT

The animal study was reviewed and approved by Protocols and Ethics Committees of the Roslin Institute and the University of Edinburgh.

AUTHOR CONTRIBUTIONS

LF, AM, RH and JO'D performed experiments and primary data analysis. SB and AG performed bioinformatic analysis of RNA and DNA sequence, respectively. AP and KB performed genetic analysis. KS performed network analysis. DH and KS analyzed the data and wrote the manuscript, with editing input from SB, AG, AP, and KB. DH conceived and funded the project.

ACKNOWLEDGMENTS

DH and KS are supported by The Mater Foundation. This work was supported in part by a Biotechnology and Biological Sciences Research Council (BBSRC) project grant (BB/M011925/1) to DH and institute strategic program grants 'Farm Animal Genomics' (BBS/E/D/20211550) and 'Transcriptomes, Networks and Systems' (BBS/E/D/20211552). RNAseq data was generate by Edinburgh Genomics which is partly supported through core grants from the BBSRC (BB/J004243/1), Natural Environmental Research Council (R8/H10/56), and Medical Research Council (MR/K001744/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01032/full#supplementary-material

REFERENCES

Ahmed-Hassan, H., Abdul-Cader, M. S., Sabry, M. A., Hamza, E., and Abdul-Careem, M. F. (2018). Toll-like receptor (TLR)4 signalling induces myeloid differentiation primary response gene (MYD) 88 independent pathway in avian species leading to type I interferon production and antiviral response. *Virus Res.* 256, 107–116. doi: 10.1016/j.virusres.2018.08.008

Baillie, J. K., Arner, E., Daub, C., De Hoon, M., Itoh, M., Kawaji, H., et al. (2017). Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genet.* 13, e1006641. doi: 10.1371/journal.pgen.1006641

Balic, A., Garcia-Morales, C., Vervelde, L., Gilhooley, H., Sherman, A., Garceau, V., et al. (2014). Visualisation of chicken macrophages using transgenic reporter genes: insights into the development of the avian macrophage lineage. *Development* 141, 3255–3265. doi: 10.1242/dev.105593

Bastepe, M. (2018). GNAS mutations and heterotopic ossification. *Bone* 109, 80–85. doi: 10.1016/j.bone.2017.09.002

Belotti, D., Capelli, C., Resovi, A., Introna, M., and Taraboletti, G. (2016). Thrombospondin-1 promotes mesenchymal stromal cell functions via TGFbeta and in cooperation with PDGF. *Matrix Biol.* 55, 106–116. doi: 10.1016/j.matbio.2016.03.003

Bilkova, B., Bainova, Z., Janda, J., Zita, L., and Vinkler, M. (2017). Different breeds, different blood: cytometric analysis of whole blood cellular composition in chicken breeds. *Vet. Immunol. Immunopathol.* 188, 71–77. doi: 10.1016/j.vetimm.2017.05.001

Biscarini, F., Bovenhuis, H., van Arendonk, J. A., Parmentier, H. K., Jungerius, A. P., and van der Poel, J. J. (2010). Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet.* 41, 26–38. doi: 10.1111/j.1365-2052.2009.01960.x

Boschiero, C., Moreira, G. C. M., Gheyas, A. A., Godoy, T. F., Gasparin, G., Mariani, P. D. S. C., et al. (2018). Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. *BMC Genomics* 19, 83. doi: 10.1186/s12864-018-4444-0

Boulton, K., Nolan, M. J., Wu, Z., Psifidi, A., Riggio, V., Harman, K., et al. (2018a). Phenotypic and genetic variation in the response of chickens to Eimeria tenella induced coccidiosis. *Genet. Sel. Evol.* 50, 63. doi: 10.1186/s12711-018-0433-7

Boulton, K., Nolan, M. J., Wu, Z., Riggio, V., Matika, O., Harman, K., et al. (2018b). Dissecting the genomic architecture of resistance to Eimeria maxima parasitism in the chicken. *Front. Genet.* 9, 528. doi: 10.3389/fgene.2018.00528

Boyd, A. C., Peroval, M. Y., Hammond, J. A., Prickett, M. D., Young, J. R., and Smith, A. L. (2012). TLR15 is unique to avian and reptilian lineages and recognizes a yeast-derived agonist. *J. Immunol.* 189, 4930–4938. doi: 10.4049/jimmunol.1101790

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Burkhardt, N. B., Roll, S., Staudt, A., Elleder, D., Hartle, S., Costa, T., et al. (2019). The long pentraxin PTX3 is of major importance among acute phase proteins in chickens. *Front. Immunol.* 10, 124. doi: 10.3389/fimmu.2019.00124

Bush, S. J., Freem, L., MacCallum, A. J., O'Dell, J., Wu, C., Afrasiabi, C., et al. (2018). Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken. *BMC Genomics* 19, 594. doi: 10.1186/s12864-018-4972-7

Bush, S. J., McCulloch, M. E. B., Summers, K. M., Hume, D. A., and Clark, E. L. (2017). Integration of quantitated expression estimates from polyA-selected and rRNA-depleted RNA-seq libraries. *BMC Bioinformatics* 18, 301. doi: 10.1186/s12859-017-1714-9

Cairns, D. M., Pignolo, R. J., Uchimura, T., Brennan, T. A., Lindborg, C. M., Xu, M., et al. (2013). Somitic disruption of GNAS in chick embryos mimics progressive osseous heteroplasia. *J. Clin. Invest.* 123, 3624–3633. doi: 10.1172/JCI69746

Campo, J. L., and Davila, S. G. (2002). Estimation of heritability for heterophil: lymphocyte ratio in chickens by restricted maximum likelihood. Effects of age, sex, and crossing. *Poult. Sci.* 81, 1448–1453. doi: 10.1093/ps/81.10.1448

Campos, R. L., Nones, K., Ledur, M. C., Moura, A. S., Pinto, L. F., Ambo, M., et al. (2009). Quantitative trait loci associated with fatness in a broiler-layer cross. *Anim. Genet.* 40, 729–736. doi: 10.1111/j.1365-2052.2009.01910.x

Chang, M. K., Raggatt, L. J., Alexander, K. A., Kuliwaba, J. S., Fazzalari, N. L., Schroder, K., et al. (2008). Osteal tissue macrophages are intercalated throughout human and mouse bone lining tissues and regulate osteoblast function in vitro and in vivo. *J. Immunol.* 181, 1232–1244. doi: 10.4049/jimmunol.181.2.1232

Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796. doi: 10.1038/nrg2664

Chaudhari, A. A., Kim, W. H., and Lillehoj, H. S. (2018). Interleukin-4 (IL-4) may regulate alternative activation of macrophage-like cells in chickens: a sequential study using novel and specific neutralizing monoclonal antibodies against chicken IL-4. *Vet. Immunol. Immunopathol.* 205, 72–82. doi: 10.1016/j.vetimm.2018.10.011

Chen, L. R., Suyemoto, M. M., Sarsour, A. H., Cordova, H. A., Oviedo-Rondon, E. O., Wineland, M., et al. (2019). Temporal characterization of wooden breast myopathy ("woody breast") severity and correlation with growth rate and lymphocytic phlebitis in three commercial broiler strains and a random-bred broiler strain. *Avian Pathol.* 48, 319–328. doi: 10.1080/03079457.2019.1598541

Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 14, 471–482. doi: 10.1038/nrg3503

Cuperus, T., van Dijk, A., Dwars, R. M., and Haagsman, H. P. (2016). Localization and developmental expression of two chicken host defense peptides: cathelicidin-2 and avian beta-defensin 9. *Dev. Comp. Immunol.* 61, 48–59. doi: 10.1016/j.dci.2016.03.008

Dai, X. M., Ryan, G. R., Hapel, A. J., Dominguez, M. G., Russell, R. G., Kapp, S., et al. (2002). Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood* 99, 111–120. doi: 10.1182/blood.V99.1.111

Erzurumluoglu, A. M., Shihab, H. A., Rodriguez, S., Gaunt, T. R., and Day, I. N. (2016). Importance of genetic studies in consanguineous populations for the characterization of novel human gene functions. *Ann. Hum. Genet.* 80, 187–196. doi: 10.1111/ahg.12150

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343, 1246949. doi: 10.1126/science.1246949

Finnberg, N., Gruber, J. J., Fei, P., Rudolph, D., Bric, A., Kim, S. H., et al. (2005). DR5 knockout mice are compromised in radiation-induced apoptosis. *Mol. Cell Biol.* 25, 2000–2013. doi: 10.1128/MCB.25.5.2000-2013.2005

Garceau, V., Balic, A., Garcia-Morales, C., Sauter, K. A., McGrew, M. J., Smith, J., et al. (2015). The development and maintenance of the mononuclear phagocyte system of the chick is controlled by signals from the macrophage colony-stimulating factor receptor. *BMC Biol.* 13, 12. doi: 10.1186/s12915-015-0121-9

Garceau, V., Smith, J., Paton, I. R., Davey, M., Fares, M. A., Sester, D. P., et al. (2010). Pivotal advance: avian colony-stimulating factor 1 (CSF-1), interleukin-34 (IL-34), and CSF-1 receptor genes and gene products. *J. Leukoc. Biol.* 87, 753–764. doi: 10.1189/jlb.0909624

Garcia-Morales, C., Nandi, S., Zhao, D., Sauter, K. A., Vervelde, L., McBride, D., et al. (2015). Cell-autonomous sex differences in gene expression in chicken bone marrow-derived macrophages. *J. Immunol.* 194, 2338–2344. doi: 10.4049/jimmunol.1401982

Gheyas, A. A., Boschiero, C., Eory, L., Ralph, H., Kuo, R., Woolliams, J. A., et al. (2015). Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Res.* 22, 205–217. doi: 10.1093/dnares/dsv005

Godoy, T. F., Moreira, G. C., Boschiero, C., Gheyas, A. A., Gasparin, G., Paduan, M., et al. (2015). SNP and INDEL detection in a QTL region on chicken chromosome 2 associated with muscle deposition. *Anim. Genet.* 46, 158–163. doi: 10.1111/age.12271

Goossens, K. E., Ward, A. C., Lowenthal, J. W., and Bean, A. G. (2013). Chicken interferons, their receptors and interferon-stimulated genes. *Dev. Comp. Immunol.* 41, 370–376. doi: 10.1016/j.dci.2013.05.020

Gow, D. J., Sester, D. P., and Hume, D. A. (2010). CSF-1, IGF-1, and the control of postnatal growth and development. *J. Leukoc. Biol.* 88, 475–481. doi: 10.1189/jlb.0310158

GTEx Project (2017). Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* 49, 1664–1670. doi: 10.1038/ng.3969

Hocking, P. M., Morrice, D. M., Law, A. S., and Burt, D. W. (2012). Many quantitative trait loci for feather growth in an F(2) broiler x layer cross collocate with body weight loci. *Br. Poult. Sci.* 53, 162–167. doi: 10.1080/00071668.2012.668613

Hu, T., Wu, Z., Bush, S. J., Freem, L., Vervelde, L., Summers, K. M., et al. (2019). Characterization of subpopulations of chicken mononuclear phagocytes that express TIM4 and CSF1R. *J. Immunol.* 202, 1186–1199. doi: 10.4049/jimmunol.1800504

Hume, D. A., Gutowska-Ding, M. W., Garcia-Morales, C., Kebede, A., Bamidele, O., et al. (2019). Functional evolution of the colony-stimulating factor 1 receptor (CSF1R) and its ligands in birds. *J. Leukoc. Biol.* 5. [Epub ahead of print] doi: 10.1002/JLB.6MA0519-172R

Jansson, M., and Laikre, L. (2014). Recent breeding history of dog breeds in Sweden: modest rates of inbreeding, extensive loss of genetic diversity and lack of correlation between inbreeding and health. *J. Anim. Breed. Genet.* 131, 153–162. doi: 10.1111/jbg.12060

Jenko, J., McClure, M. C., Matthews, D., McClure, J., Johnsson, M., Gorjanc, G., et al. (2019). Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle. *Genet. Sel. Evol.* 51, 9. doi: 10.1186/s12711-019-0452-z

Johnsson, M., Henriksen, R., Hoglund, A., Fogelholm, J., Jensen, P., and Wright, D. (2018). Genetical genomics of growth in a chicken model. *BMC Genomics* 19, 72. doi: 10.1186/s12864-018-4441-3

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14, 59. doi: 10.1186/1471-2164-14-59

Langlais, D., Barreiro, L. B., and Gros, P. (2016). The macrophage IRF8/IRF1 regulome is required for protection against infections and is associated with chronic inflammation. *J Exp. Med.* 213, 585–603. doi: 10.1084/jem.20151764

Lecoin, L., Sii-Felice, K., Pouponnot, C., Eychene, A., and Felder-Schmittbuhl, M. P. (2004). Comparison of maf gene expression patterns during chick embryo development. *Gene Expr. Patterns* 4, 35–46. doi: 10.1016/S1567-133X(03)00152-2

Lee, J. Y., Lim, W., Jo, G., Bazer, F. W., and Song, G. (2015). Estrogen regulation of phosphoserine phosphatase during regression and recrudescence of female reproductive organs. *Gen. Comp. Endocrinol.* 214, 40–46. doi: 10.1016/j.ygcen.2015.03.003

Lee, J., Mun, S., Kim, D. H., Cho, C. S., Oh, D. Y., and Han, K. (2017). Chicken (Gallus gallus) endogenous retrovirus generates genomic variations in the chicken genome. *Mob. DNA* 8, 2. doi: 10.1186/s13100-016-0085-5

Liu, C. F., Samsa, W. E., Zhou, G., and Lefebvre, V. (2017). Transcriptional control of chondrocyte specification and differentiation. *Semin. Cell Dev Biol.* 62, 34–49. doi: 10.1016/j.semcdb.2016.10.004

Liu, Y., Cheng, Y., Shan, W., Ma, J., Wang, H., Sun, J., et al. (2018). Chicken interferon regulatory factor 1 (IRF1) involved in antiviral innate immunity via regulating IFN-beta production. *Dev. Comp. Immunol.* 88, 77–82. doi: 10.1016/j.dci.2018.07.003

Lizio, M., Deviatiiarov, R., Nagai, H., Galan, L., Arner, E., Itoh, M., et al. (2017). Systematic analysis of transcription start sites in avian development. *PLoS Biol.* 15, e2002887. doi: 10.1371/journal.pbio.2002887

Mills, E. L., Ryan, D. G., Prag, H. A., Dikovskaya, D., Menon, D., Zaslona, Z., et al. (2018). Itaconate is an anti-inflammatory metabolite that activates Nrf2 via alkylation of KEAP1. *Nature* 556, 113–117. doi: 10.1038/nature25986

Minias, P. (2019). Evolution of heterophil/lymphocyte ratios in response to ecological and life-history traits: a comparative analysis across the avian tree of life. *J. Anim. Ecol.* 88, 554–565. doi: 10.1111/1365-2656.12941

Moreira, G. C., Godoy, T. F., Boschiero, C., Gheyas, A., Gasparin, G., Andrade, S. C., et al. (2015). Variant discovery in a QTL region on chromosome 3 associated with fatness in chickens. *Anim. Genet.* 46, 141–147. doi: 10.1111/age.12263

Parker, A., and Kaufman, J. (2017). What chickens might tell us about the MHC class II system. *Curr. Opin. Immunol.* 46, 23–29. doi: 10.1016/j.coi.2017.03.013

Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.* 11, 533–538. doi: 10.1038/nrg2815

Pengelly, R. J., Gheyas, A. A., Kuo, R., Mossotto, E., Seaby, E. G., Burt, D. W., et al. (2016). Commercial chicken breeds exhibit highly divergent patterns of linkage disequilibrium. *Heredity (Edinb)* 117, 375–382. doi: 10.1038/hdy.2016.47

Pettersson, M. E., and Jern, P. (2019). Whole-genome analysis of domestic chicken selection lines suggests segregating variation in erv makeups. *Genes (Basel)* 10, 162. doi: 10.3390/genes10020162

Podisi, B. K., Knott, S. A., Burt, D. W., and Hocking, P. M. (2013). Comparative analysis of quantitative trait loci for body weight, growth rate and growth curve parameters from 3 to 72 weeks of age in female chickens of a broiler-layer cross. *BMC Genet.* 14, 22. doi: 10.1186/1471-2156-14-22

Pridans, C., Raper, A., Davis, G. M., Alves, J., Sauter, K. A., Lefevre, L., et al. (2018). Pleiotropic impacts of macrophage and microglial deficiency on development in rats with targeted mutation of the Csf1r Locus. *J. Immunol.* 201, 2683–2699. doi: 10.4049/jimmunol.1701783

Psifidi, A., Russell, K. M., Matika, O., Sanchez-Molano, E., Wigley, P., Fulton, J. E., et al. (2018). The genomic architecture of fowl typhoid resistance in commercial layers. *Front. Genet.* 9, 519. doi: 10.3389/fgene.2018.00519

Qanbari, S., Rubin, C. J., Maqbool, K., Weigend, S., Weigend, A., Geibel, J., et al. (2019). Genetics of adaptation in modern chicken. *PLoS Genet.* 15, e1007989. doi: 10.1371/journal.pgen.1007989

Qiu, L., Chang, G., Li, Z., Bi, Y., Liu, X., and Chen, G. (2018). Comprehensive transcriptome analysis reveals competing endogenous RNA networks during avian leukosis virus, subgroup j-induced tumorigenesis in chickens. *Front. Physiol.* 9, 996. doi: 10.3389/fphys.2018.00996

Reading, P. C., Bozza, S., Gilbertson, B., Tate, M., Moretti, S., Job, E. R., et al. (2008). Antiviral activity of the long chain pentraxin PTX3 against influenza viruses. *J. Immunol.* 180, 3391–3398. doi: 10.4049/jimmunol.180.5.3391

Roy, S., Guler, R., Parihar, S. P., Schmeier, S., Kaczkowski, B., Nishimura, H., et al. (2015). Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection. *J. Immunol.* 194, 6035–6044. doi: 10.4049/jimmunol.1402521

Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi: 10.1038/nature08832

Rychlik, I., Elsheimer-Matulova, M., and Kyrova, K. (2014). Gene expression in the chicken caecum in response to infections with non-typhoid Salmonella. *Vet. Res.* 45, 119. doi: 10.1186/s13567-014-0119-2

Saleheen, D., Natarajan, P., Armean, I. M., Zhao, W., Rasheed, A., Khetarpal, S. A., et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* 544, 235–239. doi: 10.1038/nature22034

Salomonsen, J., Sorensen, M. R., Marston, D. A., Rogers, S. L., Collen, T., van Hateren, A., et al. (2005). Two CD1 genes map to the chicken MHC,

indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8668–8673. doi: 10.1073/pnas.0409213102

Sheikh, F., Dickensheets, H., Gamero, A. M., Vogel, S. N., and Donnelly, R. P. (2014). An essential role for IFN-beta in the induction of IFN-stimulated gene expression by LPS in macrophages. *J. Leukoc. Biol.* 96, 591–600. doi: 10.1189/jlb.2A0414-191R

Smith, J., Sadeyen, J. R., Paton, I. R., Hocking, P. M., Salmon, N., Fife, M., et al. (2011). Systems analysis of immune responses in Marek's disease virus-infected chickens identifies a gene involved in susceptibility and highlights a possible novel pathogenicity mechanism. *J. Virol.* 85, 11146–11158. doi: 10.1128/JVI.05499-11

Staines, K., Hunt, L. G., Young, J. R., and Butter, C. (2014). Evolution of an expanded mannose receptor gene family. *PLoS One* 12, e110330. doi: 10.1371/journal.pone.0110330

Stainton, J. J., Charlesworth, B., Haley, C. S., Kranis, A., Watson, K., and Wiener, P. (2017). Use of high-density SNP data to identify patterns of diversity and signatures of selection in broiler chickens. *J. Anim. Breed. Genet.* 134, 87–97. doi: 10.1111/jbg.12228

Swaggerty, C. L., Callaway, T. R., Kogut, M. H., Piva, A., and Grilli, E. (2019). Modulation of the immune response to improve health and reduce foodborne pathogens in poultry. *Microorganisms* 7, 65. doi: 10.3390/microorganisms7030065

Swaggerty, C. L., Kogut, M. H., Ferro, P. J., Rothwell, L., Pevzner, I. Y., and Kaiser, P. (2004). Differential cytokine mRNA expression in heterophils isolated from Salmonella-resistant and -susceptible chickens. *Immunology* 113, 139–148. doi: 10.1111/j.1365-2567.2004.01939.x

Swaggerty, C. L., McReynolds, J. L., Byrd, J. A., Pevzner, I. Y., Duke, S. E., Genovese, K. J., et al. (2016). Selection for pro-inflammatory mediators produces chickens more resistant to Clostridium perfringens-induced necrotic enteritis. *Poult. Sci.* 95, 370–374. doi: 10.3382/ps/pev348

Swaggerty, C. L., Pevzner, I. Y., He, H., Genovese, K. J., and Kogut, M. H. (2017). Selection for pro-inflammatory mediators produces chickens more resistant to Campylobacter jejuni. *Poult. Sci.* 96, 1623–1627. doi: 10.3382/ps/pew465

Swaggerty, C. L., Pevzner, I. Y., Kaiser, P., and Kogut, M. H. (2008). Profiling pro-inflammatory cytokine and chemokine mRNA expression levels as a novel method for selection of increased innate immune responsiveness. *Vet. Immunol. Immunopathol.* 126, 35–42. doi: 10.1016/j.vetimm.2008.06.005

Swiderska, Z., Smidova, A., Buchtova, L., Bryjova, A., Fabianova, A., Munclinger, P., et al. (2018). Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds. *Sci Rep.* 8, 17878. doi: 10.1038/s41598-018-36226-1

Thaler, J. P., Koo, S. J., Kania, A., Lettieri, K., Andrews, S., Cox, C., et al. (2004). A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. *Neuron* 41, 337–350. doi: 10.1016/S0896-6273(04)00011-X

Tran, T., Beaumont, C., Salmon, N., Fife, M., Kaiser, P., Le Bihan-Duval, E., et al. (2012). A maximum likelihood QTL analysis reveals common genome regions controlling resistance to Salmonella colonization and carrier-state. *BMC Genomics* 13, 198. doi: 10.1186/1471-2164-13-198

Wu, Z., Hu, T., Rothwell, L., Vervelde, L., Kaiser, P., Boulton, K., et al. (2016). Analysis of the function of IL-10 in chickens using specific neutralising antibodies and a sensitive capture ELISA. *Dev. Comp. Immunol.* 63, 206–212. doi: 10.1016/j.dci.2016.04.016

Young, R., Bush, S. J., Lefevre, L., McCulloch, M. E. B., Lisowski, Z. M., Muriuki, C., et al. (2018). Species-specific transcriptional regulation of genes involved in nitric oxide production and arginine metabolism in macrophages. *Immunohorizons* 2, 27–37. doi: 10.4049/immunohorizons.1700073

Yue, X., Favot, P., Dunn, T. L., Cassady, A. I., and Hume, D. A. (1993). Expression of mRNA encoding the macrophage colony-stimulating factor receptor (c-fms) is controlled by a constitutive promoter and tissue-specific transcription elongation. *Mol. Cell Biol.* 13, 3191–3201. doi: 10.1128/MCB.13.6.3191

Yukata, K., Matsui, Y., Shukunami, C., Takimoto, A., Goto, T., Nishizaki, Y., et al. (2008). Altered fracture callus formation in chondromodulin-I deficient mice. *Bone* 43, 1047–1056. doi: 10.1016/j.bone.2008.08.111

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487. doi: 10.1038/ng.3538

Zhuo, Z., Lamont, S. J., and Abasht, B. (2017). RNA-seq analyses identify frequent allele specific expression and no evidence of genomic imprinting in specific embryonic tissues of chicken. *Sci Rep.* 7, 11944. doi: 10.1038/s41598-017-12179-9

Zhuo, Z., Lamont, S. J., and Abasht, B. (2019). RNA-seq analyses identify additivity as the predominant gene expression pattern in f1 chicken embryonic brain and liver. *Genes (Basel)* 10, 27. doi: 10.3390/genes10010027

Zimmer, F., Harrison, P. W., Dessimoz, C., and Mank, J. E. (2016). Compensation of dosage-sensitive genes on the chicken z chromosome. *Genome Biol. Evol.* 8, 1233–1242. doi: 10.1093/gbe/evw075

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Freem, Summers, Gheyas, Psifidi, Boulton, MacCallum, Harne, O'Dell, Bush and Hume. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/fgene.2019.01175

Conjoint Analysis of SMRT- and Illumina-Based RNA-Sequencing Data of *Fenneropenaeus chinensis* Provides Insight Into Sex-Biased Expression Genes Involved in Sexual Dimorphism

Qiong Wang[1,2], *Yuying He*[1,2] *and Jian Li*[1,2]*

[1] *Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China,* [2] *Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China*

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Brittney Keel, United States Department of Agriculture, United States
Wenguang Liu, South China Sea Institute of Oceanology (CAS), China

***Correspondence:**
Jian Li
bigbird@ysfri.ac.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *14 July 2019*
Accepted: *24 October 2019*
Published: *15 November 2019*

Citation:
Wang Q, He Y and Li J (2019) Conjoint Analysis of SMRT- and Illumina-Based RNA-Sequencing Data of Fenneropenaeus chinensis Provides Insight Into Sex-Biased Expression Genes Involved in Sexual Dimorphism. Front. Genet. 10:1175. doi: 10.3389/fgene.2019.01175

Fenneropenaeus chinensis (*F. chinensis*) is one of the most commercially important cultured shrimps in China. The adult *F. chinensis* exhibit sexual dimorphism in growth and body color. In this research, we profiled the whole transcriptome of *F. chinensis* by using single molecule real-time-based full-length transcriptome sequencing. We further performed Illumina-based short reads RNA-seq on muscle and gonad of two sexes to detect the sex-biased expression genes. In muscle, we observed significantly more female-biased transcripts. With the differentially expressed transcripts (DETs) in muscle, some pathways related to the energy metabolism were enriched, which may be responsible for the difference of growth. We also digged out a pathway named porphyrin and chlorophyll metabolism. It was speculated to relevant to the difference of body color between the two sexes of shrimp. Interestingly, almost all DETs in these pathways were female-biased expression in muscle, which could explain the phenomenon of better growth performance and darker body color in female. In gonad, several pathways involved in reproduction were enriched. For instance, some female-biased DETs participated in the arachidonic acid metabolism, which was reported crucial in female reproduction. In conclusion, our studies identified abundant sex-biased expression transcripts and important pathways involved in sexual dimorphism by using the RNA-seq method. It provided a basis for future researches on the sexual dimorphism of *F. chinensis*.

Keywords: shrimp, full-length transcriptome, growth, body color, reproduction

INTRODUCTION

Fenneropenaeus chinensis (*F. chinensis*), which belongs to the family *Penaeidae* of *Crustacea*, is one of the most commercially important cultured shrimps in China. It mainly distributes in the Yellow Sea and Bohai Sea of China and west and south coast of the Korean Peninsula (Wang et al., 2017). Due to the delicious taste and rich nutrition, the *F. chinensis* is becoming more and more popular in consumers.

There are many species exhibit pronounced sexual dimorphism in nature. They usually show different colors, shapes, or body weight in different sexes, like chicken, peacock, guppy, and so on. The extensive sexual dimorphism in nature accord with the Darwin's conjecture that sexual selection is a force distinct from natural selection (Lande, 1980; Tipton, 1999). Sexual dimorphism is an extreme form of phenotypic plasticity. Studies on sexual dimorphism are significative for the wide intra-specific variations (Mank, 2017). In the process of cultivation, the color of adult female shrimps of *F. chinensis* tend to blue, while the males tend to yellow. The adult females were observed bigger in body size and heavier in body weight than males. Body weight is an important economical trait in production. Locate the genes related to growth will accelerate the process of molecular breeding for *F. chinensis*.

Gene expression plays an important role in generating the phenotypic diversity since there was limited genetic divergence in genome of organism. Most of sexual dimorphism are caused by the differential expression of genes between different sexes, which is known as sex-biased gene expression (Ellegren and Parsch, 2007; Grath and Parsch, 2016). Sex-biased genes could be classified as either male-biased or female-biased expression depending on which sex expresses higher (Grath and Parsch, 2016).

In recent years, short reads RNA sequencing (RNA-seq) technique has become an important tool in biological studies. It is powerful for uncovering the relationship between genotype and phenotype (Qian et al., 2014; Wang et al., 2018). However, the short reads, mostly 100~300 bp, bring many challenges to the transcriptome assemble. For instance, it is difficult to identify the alternative splicing with short reads; the repetitive sequence also could cause confusion in the assemble. Recently, the third-generation sequencing technology has sharply increased the length of sequencing reads (Grabherr et al., 2011). The PacBio platform even could sequence the whole molecule of mRNA (Rhoads and Au, 2015). Due to the much longer reads length, the complex sequence such as repetitive regions could be displayed within a single read. It could achieve the full-length (FL) sequence of transcripts and identify full coding sequences and multiple encoded isoforms (Weirather et al., 2017).

The FL transcriptome sequencing technology has prompt the overall annotation of the transcriptome and the subsequent studies in many species, such as fission yeast (Kuang et al., 2017), zebrafish (Nudelman et al., 2018), and mouse (Tardaguila et al., 2018). Especially for the non-reference genome organisms, the FL transcriptome sequencing made it possible to fully characterize the novel transcript (Grabherr et al., 2011). For example, it provides insight into the adaptive divergent function in extreme metabolism of the ruby-throated hummingbird, a non-reference species (Workman et al., 2018). In aquaculture, this technology has been applied to some analysis of important characters. FL transcriptome sequencing on pacific abalone characterized the transcriptome information for female and male individuals, and identified some sex-specific isoforms (Kim et al., 2017). For Pacific white shrimp *Litopenaeus vannamei*, a species belongs to the same genus of *Penaeus* with *F. chinensis*, transcript expression profiles survey provided insight into the immune mechanism of shrimps (Zhang et al., 2019).

Due to the abundant of repetitive sequence and high heterozygosity (Gao and Kong, 2005; Wang et al., 2008), the genome of *F. chinensis* has not been completely sequenced yet. To make a reference in this research, we used a fast growth cultured breed of *F. chinensis*, "Huanghai No. 1," which was raised by continuous selection of several generations, to profile the whole transcriptome of *F. chinensis*. We further performed short reads RNA-seq on muscle and gonad of two sexes of shrimps to detect the sex-biased expression genes.

RESULTS

Expression Profiles Delineated by the Full-Length Transcriptome Sequencing

We obtained 24.81 Gb single molecule real-time (SMRT) clean data in total. Circular consensus (CCS) sequences were extracted from the original sequence according to the condition of full passes> = 1 and the sequence accuracy > 0.90. A total of 473,469 CCS reads were extracted (**Table 1**). Among them 382,500 were full length reads non-chimeric (FLNC). The FLNC sequences were clustered and we obtained 17,470 consensus isoforms with mean length of 2,191 (**Supplementary Figure S1**). After polished, 17,279 high-quality consensus isoforms and 190 low-quality consensus isoforms were obtained. The low-quality consensus isoforms were corrected with the Illumina RNA-seq data, and merged with the high-quality FL consensus isoforms. Isoforms with high identity (>0.99) were removed redundancy. Finally, we obtained 10,795 high-quality non-redundant FL transcripts with mean length of 2,315 bp. The completeness of the non-redundant FL transcripts were assessed by BUSCO (Simao et al., 2015), and result showed that the percent of the complete transcripts identified in our project was more than 65% (**Supplementary Figure S2**). These 10,795 non-redundant FL transcripts were regarded as reference transcriptome in the following analysis.

The precursor of mRNA (pre-mRNA) has a variety of splicing types. Different exons are selected to produce different mature mRNAs, which was called alternative splicing (AS). The FL sequences were pairwise compared, and 162 AS events were detected (**Supplementary Table S1**). Simple sequence repeats (SSR) are short (1~6 bp) tandemly repeated DNA sequences. It is also known as microsatellites. In this study, we totally identified 10,941 SSR (**Supplementary Table S2**). Most of the them were mono-nucleotide repeats (**Supplementary Figure S3**). There were 10,238 coding sequences (CDS) predicted in all and 8,231 of them possessed complete open reading frames (ORF)

TABLE 1 | Summary of full-length transcriptome sequencing production.

Type	Count
Circular consensus (CCS) reads	473,469
Full length reads non-chimeric (FLNC)	382,500
FL consensus isoforms	17,470
High-quality FL transcripts	17,279
Non-redundancy high-quality FL transcripts	10,795
Alternative splicing	162
Simple sequence repeats (SSR)	10,941
coding sequences (CDS)	8,231
lncRNA	823
Annotation	9,177

(**Supplementary Table S3**). Four approaches (CPC/CNCI/CPAT/Pfam) were used to predicted long non-coding RNA (lncRNA), and 823 lncRNA were identified by all four methods consistently (**Supplementary Figure S4**). The function of 9,177 high quality FL transcripts were annotated by conjoint analysis of a series of annotation databases (**Supplementary Table S4**).

Illumina-Based Ribonucleic Acid Sequencing Data Displayed the Expression Pattern of Each Transcript

Twelve samples, including three muscles and three gonads of both male and female shrimps were sequenced. We obtained a total of 88.99 Gb clean data (**Supplementary Table S5**). After mapping with the high-quality non-redundant FL transcripts, expression level of each transcript was quantified. The principal component analysis (PCA) showed that the tissue is the most effected factor for the gene expression, meanwhile the factor of sex played a greater role in gonad than in muscle (**Figure 1**).

The median of fragments per kilobase of transcript per million fragments mapped (FPKM) distribution of expressed transcripts was observed higher in gonad than muscle (**Figure 1**). There were 131 male-biased and 689 female-biased transcripts in muscle (**Figure 1** and **Supplementary Table S6**), while in gonad the number of male-biased transcripts was 473 and female-biased transcripts was 518 (**Figure 1** and **Supplementary Table S7**). This result indicated that in muscle, significantly more transcripts expressed higher in female than male.

For the 162 AS events, we checked their expression in the two tissues. However, no significantly differential expression between different sexes was detected.

Sex-Biased Expression Transcripts in Muscle Provide Some Clues for Study of Sexual Dimorphism

The body color of adult female shrimps of *F. chinensis* tend to blue, while the males tend to yellow (**Figure 2**). The adult females showed significantly more excellent performance in body length and body weight than males (**Figures 2B, C**). Since the two sexes shared identical genomes except for several potential sex-linked regions (Xie et al., 2008), sexual dimorphism could stem from gene expression differences between sexes.

The DETs were annotated with Gene Ontology database (**Supplementary Figures S5** and **S6**), and pathway annotation analysis helps to further interpret transcript functions (**Supplementary Figure S7**). In muscle, we observed many DETs participated in pathways related to the genetic information processing, like DNA replication, mismatch repair, nucleotide excision repair, aminoacyl-transfer RNA biosynthesis, Homologous recombination, ribosome biogenesis in eukaryotes and protein processing in endoplasmic reticulum (**Figure 3**). Besides, abundant of DETs involved in the metabolism of substances and energy, like N-glycan biosynthesis, beta-alanine metabolism, histidine metabolism, fatty acid elongation, biosynthesis of unsaturated fatty acids, purine metabolism, and pyrimidine metabolism. We also observed some DETs were enriched into the pathway of porphyrin and chlorophyll metabolism, which may be relevant to the body color of the shrimps. Furthermore, a pathway related to reproduction named progesterone-mediated oocyte maturation was digged out with our result.

We have exacted the expression information of the DETs in these pathways, and found that nearly all these transcripts were female-biased expression (**Figure 4** and **Supplementary Table S8**).

Ribonucleic Acid Sequencing of Gonad Identified Important Transcripts Action on Reproduction of *Fenneropenaeus chinensis*

The DETs in gonad mostly participated in cellular processes, like focal adhesion, signaling pathways regulating pluripotency of stem cells, and lysosome (**Figure 3**). There were also some substance metabolism process pathways, like arachidonic acid metabolism, glutathione metabolism, inositol phosphate metabolism, folate biosynthesis, glycolysis/gluconeogenesis, and other glycan degradation. Some signal transduction pathways were screened out, such as MAPK signaling pathway-fly, PI3K-Akt signaling pathway, Ras signaling pathway, cyclic adenosine 3,5-monophosphate (cAMP) signaling pathway, and extracellular matrix-receptor interaction. Furthermore, a pathway named circadian rhythm was enriched by several DETs.

Interestingly, for the pathways of Fc gamma R-mediated phagocytosis, Ras signaling pathway, cAMP signaling pathway, and choline metabolism in cancer, they shared one same DET (transcript/14,675), and only this DET enriched in these pathways. This transcript expressed in males far beyond females (**Figure 5** and **Supplementary Table S9**), which action on lipid transport and metabolism. In gonad, DETs in most of pathways were female-biased or male-biased expression irregularly, while in the pathway of arachidonic acid metabolism, all the six DETs were female-biased (**Figure 5**).

DISCUSSION

In this study, we obtained 10,795 high-quality FL transcripts, which was relatively less than the similar work on *L. vannamei* (Zhang et al., 2019). Besides the species difference, it could be attribute to the strict parameters setting in the process of FL

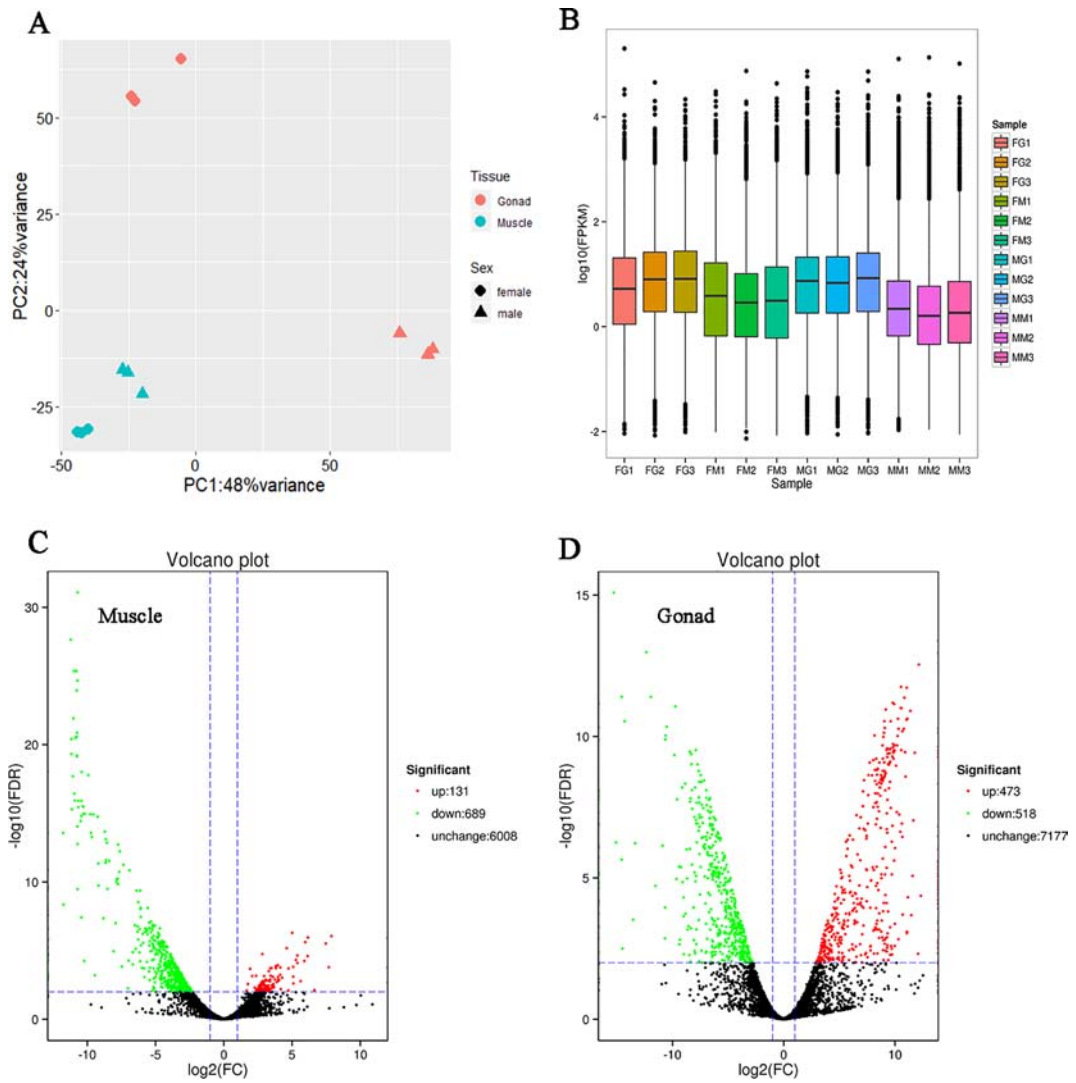


FIGURE 1 | Expression profile reflected by the short reads RNA sequencing data. **(A)** Principal component analysis of the 12 samples. Each point represents one sample, with shape indicating sex, and color indicating tissue. **(B)** Boxplot of the fragments per kilobase of transcript per million fragments mapped (FPKM) distribution of each sample. The abscissa represents different samples. The first letter of the sample name represents sex (the "F" means female, and "M" means male). The second letter of the sample name represents tissue (the "G" means gonad, and "M" means muscle). The number of the sample name represents the different individuals in the same group. The ordinate represents the logarithm of the sample expression FPKM. The graph measures the expression level of each sample from the perspective of the overall dispersion of expression quantity. The last two charts were expression volcano plot of differentially expressed transcripts (DETs) in muscle **(C)** and gonad **(D)**. Each point presents a transcript. The abscissa represents the logarithm of the expression fold change of male relative to female. A larger absolute value indicating a larger expression difference between the male and female. The ordinate represents the negative logarithm of the statistical significance of the expression difference. The larger value indicating the more significant expression difference between male and female, and the better reliability of the screened DETs. The red dots represent up-regulated DETs, the green dots represent down-regulated DETs, and the black dots represent non-DETs.

transcript cluster. In order to obtain high-quality consensus isoforms, there is possibility that multi-copy sequences of a same transcript were divided into different clusters, which inevitably resulting in redundant sequences. At the same time, degradation of the 5' end during the sequencing also could result in different copies of a same transcript being divided into different clusters. Considering about that, we clustered the FL transcripts with a strict condition. Completeness assessment of the FL transcriptome (**Supplementary Figure S2**) proved the reliability of our result.

The median of the FPKM distribution of expressed transcripts in gonad was higher than in muscle. This reflected a higher expression abundance of transcripts in gonad to a certain extent. It also could attribute to limited genes expressing in gonad. Moreover, since the FPKM value was to normalize the expression level by eliminating the effect of the sequencing depth and gene length, more longer transcript expressed in muscle could cause this result in FPKM distribute (Wagner et al., 2012).

Some aspects of sexual dimorphism result from genes located on the sex chromosomes (Rice, 1984). For *F. chinensis*, the sex

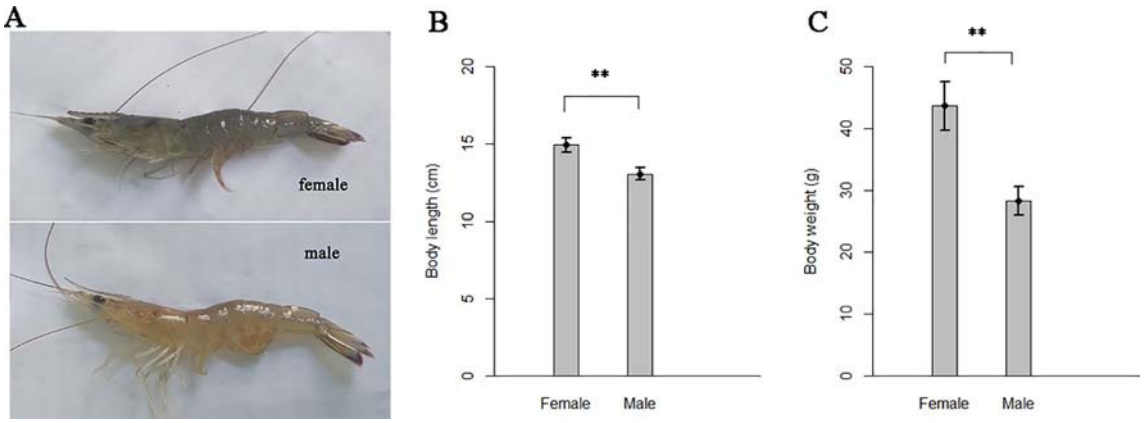


FIGURE 2 | Differences in body color and body size of the two sexes shrimps. **(A)** Body color. **(B)** Body length and **(C)** body weights at 6-months-old. Thirty female and thirty male full-sib shrimps which farmed at a same pond were measured. The line on bar represents standard deviation. "**" indicates significant differences between the two sexes with one-way analysis of variance ($P < 0.01$).

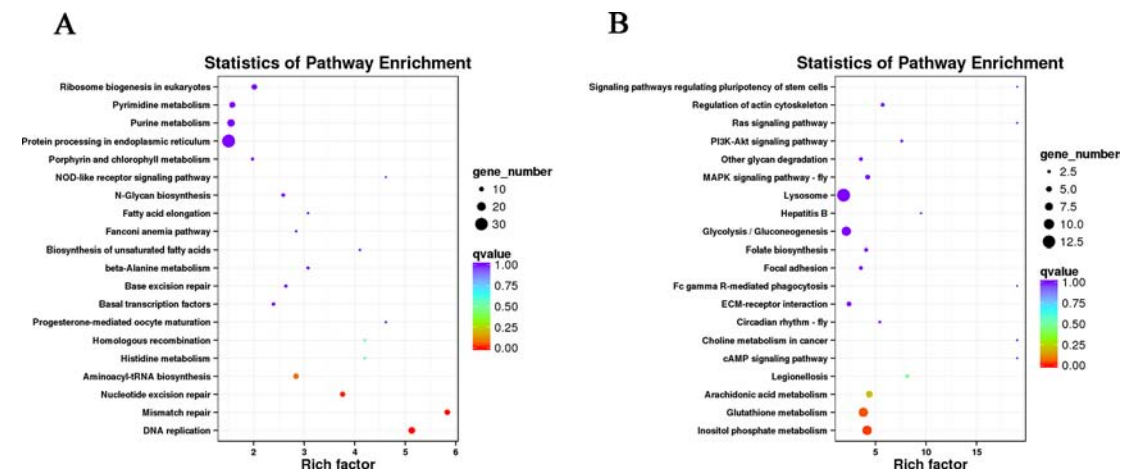


FIGURE 3 | Pathway enrichment of differentially expressed transcripts in muscle **(A)** and gonad **(B)**. Each circle represents a Kyoto Encyclopedia of Genes and Genomes pathway. The vertical axis displays the pathway name and the horizontal axis represents the enrichment factor, which indicating the ratio of the proportion of transcripts annotated into a certain pathway in differentially expressed transcripts (DETs) to the proportion of transcripts annotated into that pathway in all transcripts. A higher enrichment factor represents a more significant enrichment level of DETs in this pathway. The color of the circle represents q-value, which is the P value after correction of multiple hypothesis test. The smaller q-value indicates more reliable enrichment of DET in this pathway. The size of the circle indicates the number of transcripts enriched in the pathway.

determination and differentiation mechanism have yet to be elucidated. It remains unclear whether the ZW sex determination system (Xie et al., 2008) or some more complicated mechanism instead of simple XY or ZW system (Li et al., 2003) existing in the sex determination of *F. chinensis* (Li et al., 2012). However, many organisms lacking of sex chromosomes entirely also performed pronounced dimorphisms, and even though some species possess sex chromosomes, a majority of dimorphism controlled by genes presenting in both sexes (Bachtrog et al., 2014). Expression is one way that genes can be deployed differently. Genes show differential expression between male and female are referred to as sex-biased genes (Mank, 2017). The magnitude of sex-biased expression amplifies along with development, and reach the most manifest in adults (Mank et al., 2010; Perry et al., 2014). Therefore, we collected samples at 5-months old in this study, when they approaching the mating stage, expecting to capture more sex-biased expression genes.

The *F. chinensis* is an annual shrimp. Female shrimps migrate to warmer sea area to overwinter after mating, and swim back to original coast for oviposition. In the meantime, selective pressure forced on females. The female shrimps should store enough energy for the migration and preparation for oviposition, which could result in the bigger body size.

There were significantly more transcripts showed female-biased expression in muscle, which was speculated attributing to the better growth performance of females. With the DETs in muscle,

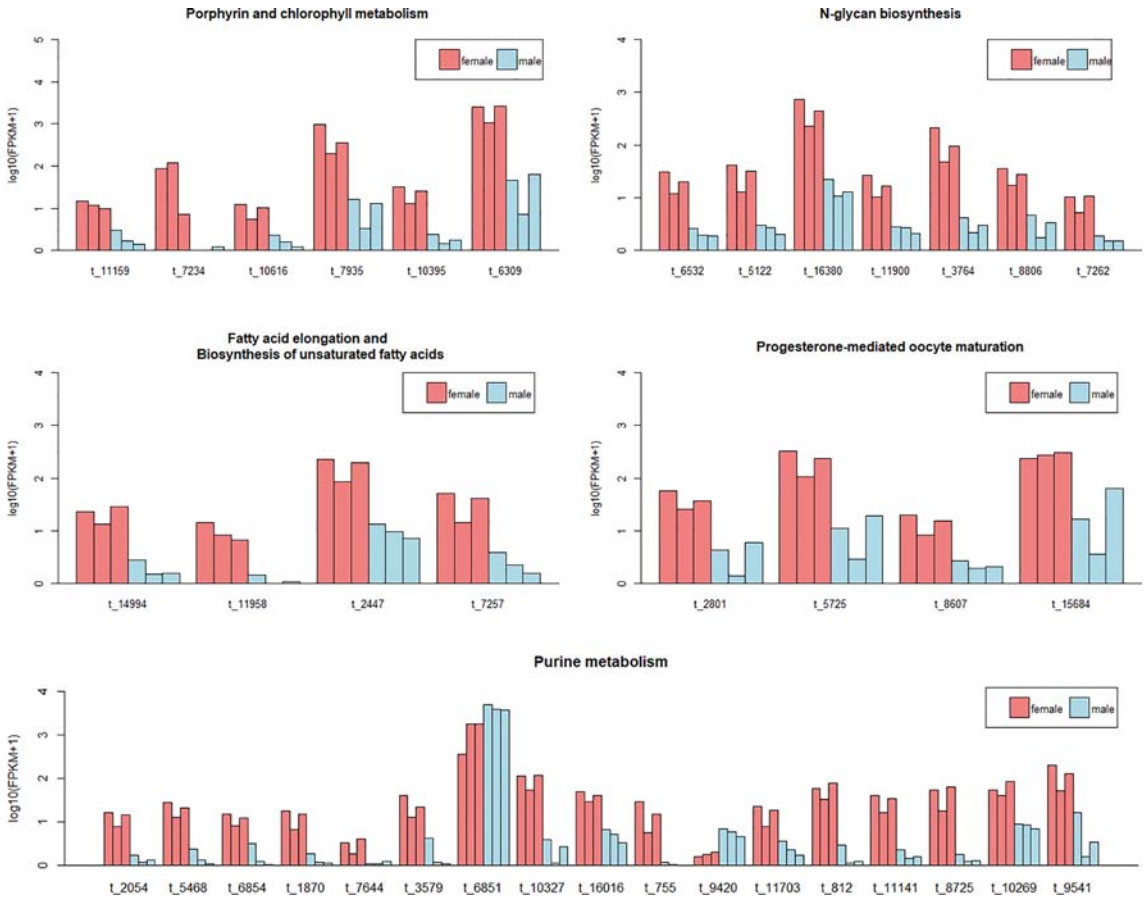


FIGURE 4 | The expression level of the differentially expressed transcripts (DETs) of muscle in five pathways. Each bar represents one sample. Different colors indicate different sexes. The abscissa represents the transcript ID. The ordinate represents the logarithm of the (FPKM+1). Since the DETs in pathways of fatty acid elongation and biosynthesis of unsaturated fatty acids were same, we combined the two pathways in one plot.

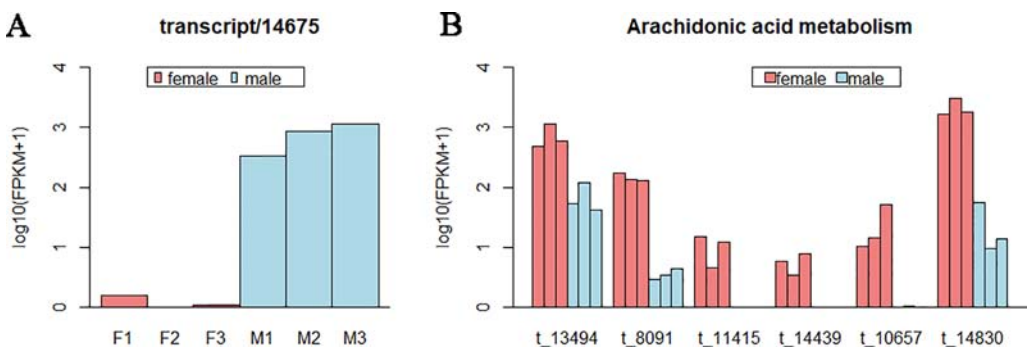


FIGURE 5 | The expression level of some differentially expressed transcripts (DETs) in gonad. **(A)** Expression level of the DET (transcript/14,675) shared by pathways of Fc gamma R-mediated phagocytosis, Ras signaling pathway, cyclic adenosine 3,5-monophosphate signaling pathway, and choline metabolism in cancer. **(B)** Expression level of DETs in the pathway of arachidonic acid metabolism. Each bar represents one sample. Different colors indicate different sexes. The abscissa represents the transcript ID. The ordinate represents the logarithm of the (FPKM+1).

some potential pathways relating to the sexual dimorphism of *F. chinensis* were unearthed. There were several pathways involved in metabolism of substances and energy. The N-glycan biosynthesis, beta-alanine metabolism, and histidine metabolism participate in protein synthesis. The fatty acid elongation and biosynthesis of unsaturated fatty acids related to fat synthesis. The purine metabolism takes part in the energy supply of organisms (Nisr and Affourtit, 2016). The metabolism of substances and energy was

considered relevant to the growth rate (Krieger, 1978; Vahl, 1984). Since most of transcripts in these pathways were female-biased, it supported the conjecture that these DETs were responsible for the fast growth of females. The fast growth always accompanied by frequent cell division and gene expression (Dayton and White, 2008), which reflected in the enrichment of DETs in the pathways of genetic information processing.

There was scarcely any research study the body color of *F. chinensis*. In this research, we caught a pathway possibly related to the pigmentation, named porphyrin and chlorophyll metabolism. Porphyrins and their derivatives are widely found in important organelles related to energy transfer in organisms (Milne et al., 2015; Galvan et al., 2016). The porphyrins show different colors when they coordinate with different metal ions. It is mainly found in heme (iron porphyrin) and hemocyanin (copper porphyrin) in animals, vitamin B12 (cobalt porphyrin), and chlorophyll (magnesium porphyrin) in plants. Interestingly, the content of a porphyrin derivative, protoporphyrin IX, was reported responsible for the brown color depth of eggshell in chicken (With, 1974). Considering about that, we proposed three hypotheses about the color difference between the two sexes: I. The hemocyanin content effect the body color; II. Other porphyrin derivatives deposited in the muscle or epidermis of female resulting in the deeper color; III. Other substances have no connection with porphyrin play roles. However, we cannot make it clear based on the current research. Further verification experiments are required to answer this question.

Although the muscle is not the reproductive tissue, there still some reproduction-related genes expressed. The pathway of progesterone-mediated maturation was digged out by the DETs in muscle. We sampled the shrimps at 5-months-old, when they approaching sexual maturity, and about to ready for the mating a month later. It is a key stage of reproduction for *F. chinensis*. The four DETs in this pathway were all female-biased, indicating an active oocyte development in female.

Unlike the muscle, DETs in gonad were irregularly female- or male-biased expression in most of pathways. Exceptionally, all the six DETs in the pathway of arachidonic acid metabolism were female-biased. The arachidonic acid (AA) is one of the initiators in prostaglandin biosynthesis (Khanapure et al., 2007), and prostaglandin could regulate reproductive function of female (Villars et al., 1985; Norberg et al., 2017). The AA was reported being largely incorporated into ovarian lipids exceeding other fatty acids (Johnson et al., 2017). It is the fatty acid precursor of an important signal molecule for crustacean reproduction (Kangpanich et al., 2016). Our result revealed that the AA was critical for the reproduction of females of *F. chinensis* at the pre-breeding stage. An appropriate feed proportion with adequate AA at this stage may be benefit for the reproduction of *F. chinensis*.

There was a transcript (transcript/14,675) expressed in male far beyond female, and participated in several important signaling pathways. This transcript was predicted to encode phospholipase D alpha 1-like, which could function on lipid transport and metabolism. The phospholipase D has been reported to play important roles on reproduction of males in other species (Vinggaard and Hansen, 1993; Lee et al., 2011; Zhang et al., 2018). This result provides a clue for research on the function of this protein in reproduction of male shrimps.

A pathway named circadian rhythm was enriched. We have known the existence of endogenous rhythms in crustacean to cope with the effect of tide, light, salinity, and so on (Naylor, 1985). The *F. chinensis* mate at a fixed time of each year, and then migrate to the south warmer area to overwinter (Kong et al., 2010). This series of behaviors were closely linked with circadian rhythm. The two DETs in this pathway was predicted to be transcribed from gene *vrille*, which was reported to drive rhythmic behavior in *Drosophila* (Gunawardhana and Hardin, 2017). We speculated that the gene *vrille* also played important role in rhythm behavior of *F. chinensis*.

In conclusion, our study profiled the transcriptome of *F. chinensis*. We further identified the DETs between two sexes which potentially responsible for the sex dimorphism in *F. chinensis*, such as growth, body color, and some reproductive-related functions. However, further researches were needed to verify the current preliminary result. Our results provided a basis for understanding the underlying molecular mechanism of sexual dimorphism in *F. chinensis*.

MATERIALS AND METHODS

Sample Collection and Handling

We picked two male and two female shrimps of "Huanghai No. 1" randomly at 5-months-old. Muscle, gonad, hepatopancreas, intestine, ganglion, heart, and sputum were collected and frozen in liquid nitrogen. Total RNA was extracted using TRIzol (Invitrogen, USA) with the standard protocols from the manufacturer. The RNA quality was assessed by NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) and agarose gel electrophoresis (AGE). Twenty-eight RNA samples (4 individuals × 7 tissues) were mixed into one pool with equal amount of nucleic acids. The mixed pool was applied to single-molecule FL transcriptome sequencing.

Another 15 female and 15 male shrimps were chosen for Illumina-based RNA-seq. We collected their gonad (female: ovary, male: testis) and muscle. Total RNA was extracted as stated above. Each five RNA samples of same sex and same tissue were mixed into one pool. The three pools of each sex were treated as biological duplicates.

To measure the body weight and body length of *F. chinensis*, we picked 30 female and 30 male full-sib shrimps of "Huanghai No. 1" randomly at 6-months-old, which were farmed at a same pond. The shrimps were measured with living body. Body length refers to the length from the base of the eyestalk to the end of the tail, when the shrimp measured as straight as possible.

Library Construction and Sequencing

We constructed FL transcriptome sequencing library of 1–6 kb complementary (cDNA) for the mixed pool sample. The library was sequenced on one SMRT Cell of Pacific Biosciences (PacBio) platform. Briefly, SMARTer™ PCR cDNA Synthesis Kit (Pacific Biosciences, Menlo Park, CA, USA) was used to generate first- and second-strand cDNA from mRNA. After a round of polymerase chain reaction (PCR) amplification and end repair,

SMRTbell™ hairpin adapters were ligated. By exonuclease digestion, we obtained a 1–6 kb cDNA library.

Twelve libraries of two tissues and two sexes (three duplicates) were constructed following the protocol of the Gene Expression Sample Prep Kit (Illumina, San Diego, CA, USA). The libraries were sequenced by Illumina NovaSeq S4 platform with paired-end (PE) 150 nt.

PacBio Long Read Processing

Raw reads were processed into error corrected reads of insert (ROIs) using Iso-Seq pipeline (Pacific Biosciences, Menlo Park, CA, USA) (Rhoads and Au, 2015) with minFullPass = 1 and minPredictedAccuracy = 0.90. FL, non-chimeric (FLNC) transcripts were determined by searching for the polyA tail signal and the 5' and 3' cDNA primers in ROIs. We used ICE (iterative clustering for error correction) to obtain FL consensus isoforms and they were further polished. Then the high-quality FL consensus isoforms were classified with the criteria post-correction accuracy above 99%. The low-quality FL consensus transcripts were corrected by our Illumina short reads RNA-seq data using the proovread software (Hackl et al., 2014), and merged with the high-quality FL consensus transcripts. Then the merged high-quality FL transcripts were removed redundancy using cd-hit (Li and Godzik, 2006) (identity > 0.99). Gene function was annotated by BLAST (Altschul et al., 1997) (version 2.2.26) based on the following databases: NR (NCBI non-redundant protein sequences) (Pruitt et al., 2007); Pfam (protein family) (Finn et al., 2014); KOG/COG/eggNOG (clusters of orthologous groups of proteins) (Tatusov et al., 2000; Koonin et al., 2004; Jensen et al., 2008); Swiss-Prot (a manually annotated and reviewed protein sequence database) (The UniProt, Consortium, 2017); KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al., 2004); GO (Gene Ontology) (Ashburner et al., 2000).

The structure analysis of the transcriptome was as follows:

Alternative Splice

We used Iso-Seq™ data directly to run all-*vs.*-all BLAST with high identity settings, BLAST alignments that met all criteria were considered products of candidate AS events: there should be two HSPs (high-scoring segment pair) larger than 1,000 bp in the alignment; the two HSPs have same forward/reverse direction, within the same alignment; one sequence should be continuous, or with a small "overlap" size (smaller than 5 bp), the other one should be distinct to show an "AS gap"; the continuous sequence should pretty much completely align to the distinct sequence; the AS Gap should larger than 100 bp and at least 100 bp away from the 3'/5' end.

Simple Sequence Repeat Detection

Simple sequence repeats (SSRs) of the transcriptome were identified using MISA (http://pgrc.ipk-gatersleben.de/misa/).

Coding Sequence Detection

Candidate coding regions within transcript sequences were identified by TransDecoder (https://github.com/TransDecoder/TransDecoder/releases) (version 5.5.0). We used the following criteria: 1) a minimum length open reading frame (ORF) is found in a transcript sequence; 2) a log-likelihood score similar to what is computed by the GeneID software (http://genome.crg.es/software/geneid/) is > 0; 3) the above coding score is greatest when the ORF is scored in the 1st reading frame as compared to scores in the other five reading frames; 4) if a candidate ORF is found fully encapsulated by the coordinates of another candidate ORF, the longer one is reported. However, a single transcript can report multiple ORFs (allowing for operons, chimeras, etc); 5) optional the putative peptide has a match to a Pfam domain above the noise cutoff score.

Long Non-Coding Ribonucleic Acid Analysis

Four computational approaches include coding potential calculator (CPC) (Kong et al., 2007), Coding-Non-Coding Index (CNCI) (Sun et al., 2013), Coding Potential Assessment Tool (CPAT) (Wang et al., 2013), and Pfam database (Finn et al., 2014) were combined to sort non-protein coding RNA candidates from putative protein-coding RNAs in the transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts with length more than 200 nt and possess more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam that have the power to distinguish the protein-coding genes from the non-coding genes.

Illumine-Based Ribonucleic Acid Sequencing Data Processing

Raw reads of FASTQ format were firstly processed through in-house Perl scripts. Briefly, clean reads were obtained by removing reads containing adapter or ploy-N and low-quality reads from raw data. At the same time, Q20, Q30, GC-content, and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

The clean reads of each RNA-seq library were aligned to the FL reference transcriptome to obtain unique mapped reads by using the tool of STAR (Dobin et al., 2013) (version 2.5.0b) with default parameters. Only reads with a perfect match or one mismatch were further analyzed and annotated based on the reference transcriptome. The read counts were adjusted by edgeR program package (Robinson et al., 2010) (version 3.22.0). Expression level of each transcript for each tissue was calculated and normalized into FPKM values by RSEM software (Li and Dewey, 2011) (version 1.2.19). The resulting FDR (false discovery rate) was adjusted using the PPDE (posterior probability of being DE) method in EBSeq package (Leng et al., 2013) (version 1.24.0). We set the conditions of FDR < 0.05 and $|\log2(\text{foldchange})| \geq 1$ as the threshold for significantly differential expression.

Functional Enrichment Analysis of DETs

GO enrichment analysis of the DETs was implemented by the R packages of GOseq (Young et al., 2010) (version 1.34.1) based on the Wallenius non-central hyper-geometric distribution, which can adjust for gene length bias in differential expression genes (DEGs). All of the transcripts of *F. chinensis* annotated in this study were used as the background data.

KEGG (Kanehisa et al., 2017) is a database resource for understanding high-level functions and utilities of the biological system (http://www.genome.jp/kegg/). We used KOBAS software (Mao et al., 2005) (version 3.0.0) to test the statistical enrichment of DETs in KEGG pathways.

DATA AVAILABILITY STATEMENT

The sample information was registered as BioProject with accession number PRJNA558194 and BioSample with accession number from SAMN12429759 to SAMN12429771. Raw sequence generated by SMRT and Illumina platform was deposited into the NCBI Sequence Read Archive (SRA) with accession number SRR9894260.

ETHICS STATEMENT

The animal study was reviewed and approved by The Animal Care and Use Committee of the Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences.

AUTHOR CONTRIBUTIONS

QW analyzed and interpreted the sequencing data and drafted the manuscript. YH participated in collecting the samples and improved the manuscript. JL conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

FUNDING

This work was supported by the China Agriculture Research System (Grant No. CARS-48), grants from the Program of Taishan Industrial Experts (Grant No. LNJY2015002), National Natural Science Foundation of China (Grant No. 31902367) and China Postdoctoral Science Foundation (Grant No. 2018M642730).

ACKNOWLEDGMENTS

We would like to appreciate our colleagues of the Key Laboratory for Sustainable Utilization of Marine Fisheries Resources of Yellow Sea Fisheries Research Institute, for their assistance on sample collection and helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01175/full#supplementary-material

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402. doi: 10.1093/nar/25.17.3389

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., et al. (2014). and XXXC. Tree of Sex: Sex determination: why so many ways of doing it? *PloS Biol.* 12 (7), e1001899. doi: 10.1371/journal.pbio.1001899

Dayton, W. R., and White, M. E. (2008). Cellular and molecular regulation of muscle growth and development in meat animals. *J. Anim. Sci.* 86 (14 Suppl), E217–E225. doi: 10.2527/jas.2007-0456

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* 8, 689. doi: 10.1038/nrg2167

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Galvan, I., Camarero, P. R., Mateo, R., and Negro, J. J. (2016). Porphyrins produce uniquely ephemeral animal colouration: a possible signal of virginity. *Sci. Rep.* 6, 39210. doi: 10.1038/srep39210

Gao, H., and Kong, J. (2005). The microsatellites and minisatellites in the genome of Fenneropenaeus chinensis. *DNA Seq.* 16 (6), 426–436. doi: 10.1080/10425170500354359

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652. doi: 10.1038/nbt.1883

Grath, S., and Parsch, J. (2016). Sex-Biased Gene Expression. *Annu. Rev. Genet.* 50 (1), 29–44. doi: 10.1146/annurev-genet-120215-035429

Gunawardhana, K. L., and Hardin, P. E. (2017). VRILLE Controls PDF neuropeptide accumulation and arborization rhythms in small ventrolateral neurons to drive rhythmic behavior in Drosophila. *Curr. Biol.* 27 (22), 3442–3453 e4. doi: 10.1016/j.cub.2017.10.010

Hackl, T., Hedrich, R., Schultz, J., and Forster, F. (2014). proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30 (21), 3004–3011. doi: 10.1093/bioinformatics/btu392

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., et al. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* 36 (Database issue), D250–D254. doi: 10.1093/nar/gkm796

Johnson, R. B., Kroeger, E. L., Reichert, W. L., Carter, C. S., and Rust, M. B. (2017). Uptake and selective partitioning of dietary lipids to ovarian and muscle tissue of maturing female coho salmon, Oncorhynchus kisutch, during secondary oocyte growth. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 208-209, 7–18. doi: 10.1016/j.cbpb.2017.03.005

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32 (Database issue), D277–D280. doi: 10.1093/nar/gkh063

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi: 10.1093/nar/gkw1092

Kangpanich, C., Pratoomyot, J., Siranonthana, N., and Senanan, W. (2016). Effects of arachidonic acid supplementation in maturation diet on female reproductive performance and larval quality of giant river prawn (Macrobrachium rosenbergii). *PeerJ* 4, e2735. doi: 10.7717/peerj.2735

Khanapure, S. P., Garvey, D. S., Janero, D. R., and Letts, L. G. (2007). Eicosanoids in inflammation: biosynthesis, pharmacology, and therapeutic frontiers. *Curr. Top. Med. Chem.* 7 (3), 311–340. doi: 10.2174/156802607779941314

Kim, M. A., Rhee, J. S., Kim, T. H., Lee, J. S., Choi, A. Y., Choi, B. S., et al. (2017). Alternative Splicing Profile and Sex-Preferential Gene Expression in the Female and Male Pacific Abalone Haliotis discus hannai. *Genes (Basel)* 8 (3), 99. doi: 10.3390/genes8030099

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35 (Web Server issue), W345–W349. doi: 10.1093/nar/gkm391

Kong, X. Y., Li, Y. L., Shi, W., and Kong, J. (2010). Genetic variation and evolutionary demography of Fenneropenaeus chinensis populations, as revealed by the analysis of mitochondrial control region sequences. *Genet. Mol. Biol.* 33 (2), 379–389. doi: 10.1590/S1415-47572010005000019

Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5 (2), R7. doi: 10.1186/gb-2004-5-2-r7

Krieger, I. (1978). Relation of specific dynamic action of food (SDA) to growth in rats. *Am. J. Clin. Nutr.* 31 (5), 764–768. doi: 10.1093/ajcn/31.5.764

Kuang, Z., Boeke, J. D., and Canzar, S. (2017). The dynamic landscape of fission yeast meiosis alternative-splice isoforms. *Genome Res.* 27 (1), 145–156. doi: 10.1101/gr.208041.116

Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34 (2), 292–305. doi: 10.1111/j.1558-5646.1980.tb04817.x

Lee, Y. J., Ahn, M. Y., Kim, H. S., Kwack, S. J., Park, K. L., Yoon, S., et al. (2011). Role of phospholipase D in regulation of testicular Leydig cell hyperplasia in Sprague-Dawley rats treated with di(2-ethylhexyl) phthalate. *Arch. Toxicol.* 85 (8), 975–985. doi: 10.1007/s00204-010-0618-5

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M., et al. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29 (8), 1035–1043. doi: 10.1093/bioinformatics/btt087

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22 (13), 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, F., Xiang, J., Zhang, X., Zhou, L., Zhang, C., and Wu, C. (2003). Gonad development characteristics and sex ratio in triploid Chinese shrimp (Fenneropenaeus chinensis). *Mar. Biotechnol. (NY)* 5 (6), 528–535. doi: 10.1007/s10126-002-0103-0

Li, S., Li, F., Wen, R., and Xiang, J. (2012). Identification and characterization of the sex-determiner transformer-2 homologue in Chinese shrimp, Fenneropenaeus chinensis. *Sex. Dev.* 6 (5), 267–278. doi: 10.1159/000341377

Mank, J. E., Nam, K., Brunstrom, B., and Ellegren, H. (2010). Ontogenetic complexity of sexual dimorphism and sex-specific selection. *Mol. Biol. Evol.* 27 (7), 1570–1578. doi: 10.1093/molbev/msq042

Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nat. Ecol. Evol.* 1 (1), 6. doi: 10.1038/s41559-016-0006

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21 (19), 3787–3793. doi: 10.1093/bioinformatics/bti430

Milne, B. F., Toker, Y., Rubio, A., and Nielsen, S. B. (2015). Unraveling the intrinsic color of chlorophyll. *Angew. Chem. Int. Ed. Engl.* 54 (7), 2170–2173. doi: 10.1002/anie.201410899

Naylor, E. (1985). Tidally rhythmic behaviour of marine animals. *Symp. Soc. Exp. Biol.* 39, 63–93. doi: 10.1016/b978-0-408-70778-7.50010-x

Nisr, R. B., and Affourtit, C. (2016). Palmitate-induced changes in energy demand cause reallocation of ATP supply in rat and human skeletal muscle cells. *Biochim. Biophys. Acta* 1857 (9), 1403–1411. doi: 10.1016/j.bbabio.2016.04.286

Norberg, B., Kleppe, L., Andersson, E., Thorsen, A., Rosenlund, G., and Hamre, K. (2017). Effects of dietary arachidonic acid on the reproductive physiology of female Atlantic cod (Gadus morhua L.). *Gen. Comp. Endocrinol.* 250, 21–35. doi: 10.1016/j.ygcen.2017.05.020

Nudelman, G., Frasca, A., Kent, B., Sadler, K. C., Sealfon, S. C., Walsh, M. J., et al. (2018). High resolution annotation of zebrafish transcriptome using long-read sequencing. *Genome Res.* 28 (9), 1415–1425. doi: 10.1101/gr.223586.117

Perry, J. C., Harrison, P. W., and Mank, J. E. (2014). The ontogeny and evolution of sex-biased gene expression in Drosophila melanogaster. *Mol. Biol. Evol.* 31 (5), 1206–1219. doi: 10.1093/molbev/msu072

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35 (Database issue), D61–D65. doi: 10.1093/nar/gkl842

Qian, X., Ba, Y., Zhuang, Q., and Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *Omics* 18 (2), 98–110. doi: 10.1089/omi.2013.0110

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinf.* 13 (5), 278–289. doi: 10.1016/j.gpb.2015.08.002

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evol.* 38 (4), 735–742. doi: 10.1111/j.1558-5646.1984.tb00346.x

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinf.* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinf.* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41 (17), e166. doi: 10.1093/nar/gkt646

Tardaguila, M., Marti, C., Pereira, C., Pardo-Palacios, F. J., Del Risco, H., Ferrell, M., et al. (2018). SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* 28 (3), 396–411 doi: 10.1101/gr.222976.117

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28 (1), 33–36. doi: 10.1093/nar/28.1.33

The UniProt, Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–D169. doi: 10.1093/nar/gkw1099

Tipton, J. A. (1999). Darwin's beautiful notion: sexual selection and the plurality of moral codes. *Hist. Philos. Life Sci.* 21 (2), 119–135.

Vahl, O. (1984). The relationship between specific dynamic action (SDA) and growth in the common starfish, Asterias rubens L. *Oecologia* 61 (1), 122–125. doi: 10.1007/BF00379097

Villars, T. A., Hale, N., and Chapnick, D. (1985). Prostaglandin-F2 alpha stimulates reproductive behavior of female paradise fish (Macropodus opercularis). *Horm. Behav.* 19 (1), 21–35. doi: 10.1016/0018-506x(85)90003-0

Vinggaard, A. M., and Hansen, H. S. (1993). Bradykinin and vasopressin activate phospholipase D in rat Leydig cells by a protein kinase C-dependent mechanism. *J. Endocrinol.* 136 (1), 119–126. doi: 10.1677/joe.0.1360119

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131 (4), 281–285. doi: 10.1007/s12064-012-0162-3

Wang, H., Li, F., Xiang, J., Zhang, C., and Yu, K. (2008). Microsatellite-centromere distances and microsatellite diversity in different ploidy classes of Chinese shrimp (Fenneropenaeus Chinensis). *Genetica* 132 (1), 43–50. doi: 10.1007/s10709-007-9147-4

Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41 (6), e74. doi: 10.1093/nar/gkt006

Wang, M., Kong, J., Meng, X., Luan, S., Luo, K., Sui, J., et al. (2017). Evaluation of genetic parameters for growth and cold tolerance traits in Fenneropenaeus chinensis juveniles. *PloS One* 12 (8), e0183801. doi: 10.1371/journal.pone.0183801

Wang, N., Wang, R., Wang, R., and Chen, S. (2018). Transcriptomics analysis revealing candidate networks and genes for the body size sexual dimorphism of Chinese tongue sole (Cynoglossus semilaevis). *Funct. Integr. Genomics* 18 (3), 327–339. doi: 10.1007/s10142-018-0595-y

Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X. J., et al. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* 6, 100. doi: 10.12688/f1000research.10571.2

With, T. K. (1974). Porphyrins in egg shells. *Biochem. J.* 137 (3), 597–598. doi: 10.1042/bj1370596

Workman, R. E., Myrka, A. M., Wong, G. W., Tseng, E., Welch, K. C. Jr, and Timp, W. (2018). Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated

hummingbird Archilochus colubris. *Gigascience* 7 (3), 1–12. doi: 10.1093/gigascience/giy009

Xie, Y., Li, F., Zhang, C., Yu, K., and Xiang, J. (2008). Synaptonemal complex analysis in spermatocytes of diploid and triploid Chinese shrimp Fenneropenaeus chinensis. *Tissue Cell* 40 (5), 343–350. doi: 10.1016/j.tice.2008.03.003

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11 (2), R14. doi: 10.1186/gb-2010-11-2-r14

Zhang, P., Qin, Y., Zheng, Y., and Zeng, W. (2018). Phospholipase D Family Member 6 Is a Surface Marker for Enrichment of Undifferentiated Spermatogonia in Prepubertal Boars. *Stem Cells Dev.* 27 (1), 55–64. doi: 10.1089/scd.2017.0140

Zhang, X., Li, G., Jiang, H., Li, L., Ma, J., Li, H., et al. (2019). Full-length transcriptome analysis of Litopenaeus vannamei reveals transcript variants involved in the innate immune system. *Fish Shellfish Immunol.* 87, 346–359. doi: 10.1016/j.fsi.2019.01.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, He and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/fgene.2019.01010

Genome-Wide Differential Expression Profiling of Ovarian circRNAs Associated With Litter Size in Pigs

*Gaoxiao Xu[1,2], Huifang Zhang[1], Xiao Li[1], Jianhong Hu[1], Gongshe Yang[1] and Shiduo Sun[1]**

[1] Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Sciences and Technologies, Northwest A&F University, Yangling, China, [2] Teaching and Research Section of Biotechnology, Nanning University, Nanning, China

Circular RNAs (circRNAs) have been emerging as an important regulator in mammalian reproduction *via* acting as miRNA sponges. However, the circRNAs in porcine ovaries related with litter size remains largely unknown. In this study, porcine ovaries with smaller or larger litter size (LLS) were subjected to high-throughput RNA sequencing. In total, 38,722 circRNAs were identified, of which 1,291 circRNAs were commonly expressed in all samples. There were 56 circRNAs significantly down-regulated and 54 circRNAs up-regulated in LLS pig (|log2 (fold change)| > 1, FDR < 0.05). Bioinformatics predicted that most of circRNAs harbored miRNA binding sites, and the expression patterns of circRNAs and their putative binding miRNAs were validated by qPCR. Moreover, the expression of circ-*TCP11*/miR-183 was significantly reversely correlated and their direct interaction was confirmed by dual-luciferase assay. Our study indicates that circRNAs may play potential effects on modulating porcine litter size.

Keywords: circRNA, ovary, litter size, pig, miRNA

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Shahin Eghbalsaied, Islamic Azad University, Isfahan, Iran
Wenguang Liu, South China Sea Institute of Oceanology (CAS), China

***Correspondence:**
Shiduo Sun
SSdsm@Tom.com

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *17 June 2019*
Accepted: *23 September 2019*
Published: *15 November 2019*

Citation:
Xu G, Zhang H, Li X, Hu J, Yang G and Sun S (2019) Genome-Wide Differential Expression Profiling of Ovarian circRNAs Associated With Litter Size in Pigs. Front. Genet. 10:1010. doi: 10.3389/fgene.2019.01010

INTRODUCTION

Increasing litter size has been a global goal for pig breeders and producers, and larger litter size plus shorter farrowing intervals are desperately expected to expand piglets per sow per year, which is the predominant force to boost economic success of sow husbandry (Zak et al., 2017; Kemp et al., 2018). Ovary is an important reproductive organ in females and goes through a series of biological processes during each estrous cycle. Sow prolificacy is tightly modulated by the complex transcriptional network involving coding and non-coding genes in ovaries (Zhang et al., 2015; Huang et al., 2016; Tang et al., 2018).

Covalently closed circular RNAs (circRNAs) are emerging as a novel class of modulators for gene expression (Li et al., 2018). Now, accumulating work has shed light on the critical roles of circRNAs in gonadal development and reproduction performance in many species (Quan and Li 2018). Next-generation sequencing has revealed that endogenous circRNAs are generally expressed in various kinds of porcine tissues in a spatio-temporally specific manner, including ovaries (Liang et al., 2017). Recent studies revealed that human ovary-derived circRNAs are involved in ovarian aging (Cai et al., 2018), thus we investigated whether circRNAs' profile differs in sows with different litter sizes.

In this study, a total of six ovaries were selected from multiparous sows with intact prolificacy records, and high-throughput sequencing technology coupled with bioinformatic tools were employed to uncover litter-size-related circRNAs, providing potential candidate loci that may be informative for future pig breeding programs.

MATERIALS AND METHODS

Ethics Statement and Sample Collection

This study was approved by the Animal Care and Use Committee in Northwest A&F University (No. 2018-019). The ovaries in our study were collected from sows 4 days after the fourth delivery picked from a commercial sow piggery in Hanshiwei Food Ltd., Co. (Dahua, Guangxi, China), which is negative for PRRSV (porcine reproductive and respiratory syndrome virus) and PCV (porcine circovirus). For RNA-seq, three ovaries were sampled from each group small and large litter size (8.48 ± 0.53 piglets/litter in small litter size and 16.19 ± 0.43 piglets/litter in large litter size). In RT-qPCR assay, more ovaries were sampled (8.78 ± 1.75 piglets/litter *vs.* 14.83 ± 1.61 piglets/litter, n = 12). All sows were slaughtered in a standard slaughterhouse in Xinyouxian Livestock Ltd., Co., (Xining, Guangxi, China), and the left ovaries were quickly taken and frozen in liquid nitrogen.

RNA-Seq Assay

The frozen ovary tissues were homogenized in TRIzol™ reagent (Invitrogen, Carlsbad, CA, USA), and each sample was quantified using ND-1000 Nanodrop (Thermo Fisher, Wilmington, DE, USA). RNA integrity number (RIN) was analyzed with Agilent 2200 (Agilent, Palo Alto, CA, USA), and RNAs with RIN >7.0 were used for RNA-seq analysis.

The total RNA samples (3 μg) were treated with Epicenter Ribo-Zero rRNA removal kit (Illumina, San Diego, CA, USA) to remove ribosomal RNA (rRNA) before cDNA library construction, and then ribosome depleted RNAs were fragmented into 150–200 nt by incubation with divalent cations at 94°C for 8 min. The cleaved RNA fragments were reverse-transcribed into first- and second-strand cDNA according to the description of TruSeq RNA LT/HT sample preparation kit (Illumina, USA). Briefly, the cDNA was treated with End-It DNA End Repair Kit to repair the ends, then modified with Klenow to add an A at the 3′ end, and finally ligated to indexed adapters. The ligated cDNA products were purified and treated with uracil DNA glycosylase to remove the second-strand cDNA. Purified first-strand cDNA was enriched by 13–16 cycles of PCR amplification. The final cDNA libraries were evaluated by Bioanalyzer 2200 (Agilent, Santa Clara, CA) and subjected to sequencing by HiSeq 2000 (Illumina, USA).

miRNA libraries were constructed by Ion Total RNA-Seq Kit v2.0 (Life Technologies), and the sizes were selected by PAGE gel and processed for miRNA sequencing.

Bioinformatic Analysis

Raw sequencing data were tested by performing FAST-QC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and evaluation metrics including quality scores, distribution of nucleotides, GC content, k-mer frequency, and others. Low-quality bases and N bases were trimmed from the reads by NGSQCToolkit (v2.3.3), and high-quality clean reads were obtained for subsequent analysis. Clean reads were mapped to the reference genome (Sscrofa 11.1 assembly, http://genome.ucsc.edu) using Hisat2 software.

CIRI was used to identify circRNAs. The alignment results (SAM format) were scanned to search paired chiastic clipping and paired-end mapping signals, as well as GT-AG splicing signals. All the sequences with junction sites were realigned to reference genome using dynamic programming algorithm to ensure the reliability of the putative circRNAs. The total number of reads spanning back-spliced junctions was used as an absolute measure of circRNA abundance.

DEseq2 was used to explore the differentially expressed mRNA among different groups, and the criteria were set as |log2 (fold change)| > 1, FDR < 0.05. EdgeRSeq was used to explore the differentially expressed circRNA and miRNA between groups, with cutoff of |log2 (fold change)| > 1, FDR < 0.05. Gene ontology (GO) function (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg) pathway of genes of target were annotated.

TABLE 1 | Sequences of primers used in this study.

Primer list	Sequence (5′→3′)	Product length (bp)
circ-*ERBIN*	F: TTCGACATCCCCAGACATCC R: CAGACATCCGAGACGGAGAA	218
circ-*SNTB2*	F: CAACAATGGAGACCCGTCCT R: TCCTTGGTGCTGTTCTGGTG	267
circ-*TCP1*	F: GGGACCTTAAACGCATTGCTA R: GGTCCAGTTTTTCAGGGTCTGT	147
circ-*KMT2A*	F: AGGAGAACGCAGGCACTTTG R: GGAGGAGGTTCACTGTTGCT	261
circ-*LOC397451*	F: GCTTGCATTGAAAACGGGTCTC R: TCCATTACAGGCAGGACAGTG	261
circ-*NUP98*	F: AGCACAGGGACCAGTCTTTTC R: AGGCTTCCAGTATTGTTGCTG	248
circ-*SENP2*	F: GGGGAAGAGCAAAGTCATGGA R: TCCGTGTGCCATTACAAGCA	294
circ-*CCDC85A*	F: CTGCTAGACTTGACCAGCGTT R: CAAATGTGGGCCAATGGTGAT	180
circ-*CCAR1*	F: CCAACATCAGCAGCCCTTGT R: TGCTGCAATCCGAGTATCCC	218
ACTB	F: GGACTTCGAGCAGGAGATGG R: AGGAAGGAGGGCTGGAAGAG	134

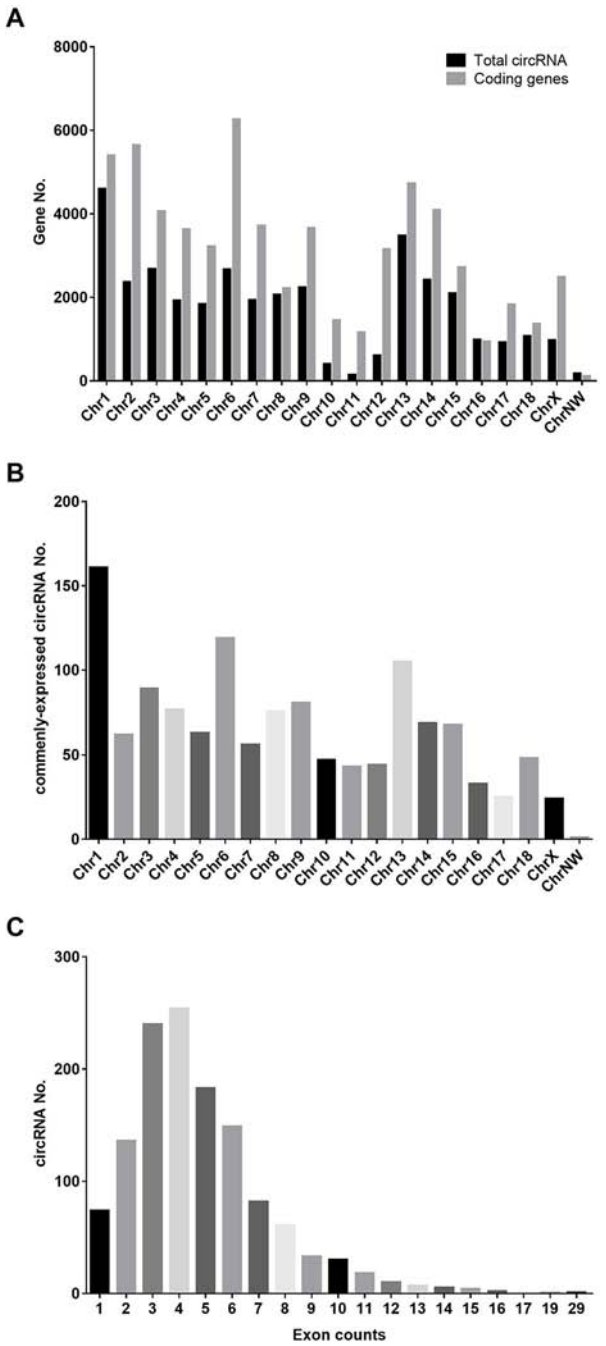


FIGURE 1 | Distribution of predicted circRNAs in our study. **(A)** The distribution of annotated mRNAs on the genome. **(B)** The distribution of commenly-expressed circRNAs on the genome. **(C)** The exon numbers of commenly-expressed circRNAs in our study.

CeRNA Network Construction

The potential miRNA–circRNA interactions were predicted by miRanda (http://miranda.org.uk/) and RNAhybrid2 (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid). The correlation between the expression levels of circRNA and miRNA was calculated with SPSS Pearson correlation assay.

RT-qPCR Verification

Total RNA was purified using TRIzol™ reagent (Invitrogen). An aliquot of 2 μg total RNA was taken from each sample and reverse transcripted by random primers (TakaRa, Otsu, Japan). For miRNA analysis, specific reverse transcription primers and procedures were used. Real-time PCR reaction (95°C 30 s, then 95°C 5 s, 60°C 30 s for 40 cycles, following 70°C 10 min for elongation) was performed in triplicate using the One-step SYBY PrimeScript RT-PCR kit (TakaRa) on a Bio-Rad iQ5™ system (Bio-Rad, Berkeley, CA, USA). The expressions of circRNA and miRNA were normalized to that of *ACTB* and *U6* small RNA, respectively. The primer sequences for qPCR are shown in **Table 1**.

Dual Luciferase Assay

A ~400 bp fragment of circRNA containing the putative miRNA binding site was synthesized by General Biosystems (Chuzhou, Anhui, China) and inserted into psiCHECK™-2 vector (Promega, Madison, WI, USA) to construct psi-circRNA plasmids. Then the psi-circRNA constructs were co-transfected with their corresponding miRNAs (RiboBio, Guangzhou, China) into 293T cells using Lipofectin™ 2000 (Thermo Fisher Scientific, Waltham, MA, USA), and the luciferase activity was detected by Dual-Luciferase® Reporter Assay System (Promega) 24 h post transfection.

Statistical Analysis

Data were processed with SPSS 19.0 software, and results were presented as mean ± SEM. Significant differences were assessed by unpaired Student's *t*-test and $p < 0.05$ was defined as statistical significance.

RESULTS

Overview of CircRNAs in Porcine Ovary

Three ovaries in each group (small litter size *vs.* large litter size) were subjected to RNA sequencing, and a total of 38,722 circRNAs were predicted, which were widely distributed across all chromosomes (**Figures 1A**, **B**). However, only 1,291 circRNAs were expressed in all samples (**Figure 1C**).

Differential CircRNA Expression Profiles in Pigs Differing Litter Size

In our study, 110 circRNAs [56 down-regulated and 54 up-regulated in larger litter size (LLS)] (**Figure 2 A**, **B**) and 20 miRNAs [11 down-regulated and 9 up-regulated in smaller litter size (SLS)] were identified by RNA-seq (**Figure 2 C, D**).

Given the high variability between samples in RNA-seq, the sample pool was expanded to 12 per group, and a total of 24 ovaries were used when confirming the differentially expressed circRNAs and miRNAs revealed above using RT-PCR. Based on the expanded sample size, RT-PCR assay uncovered that circ-*ERBN*,

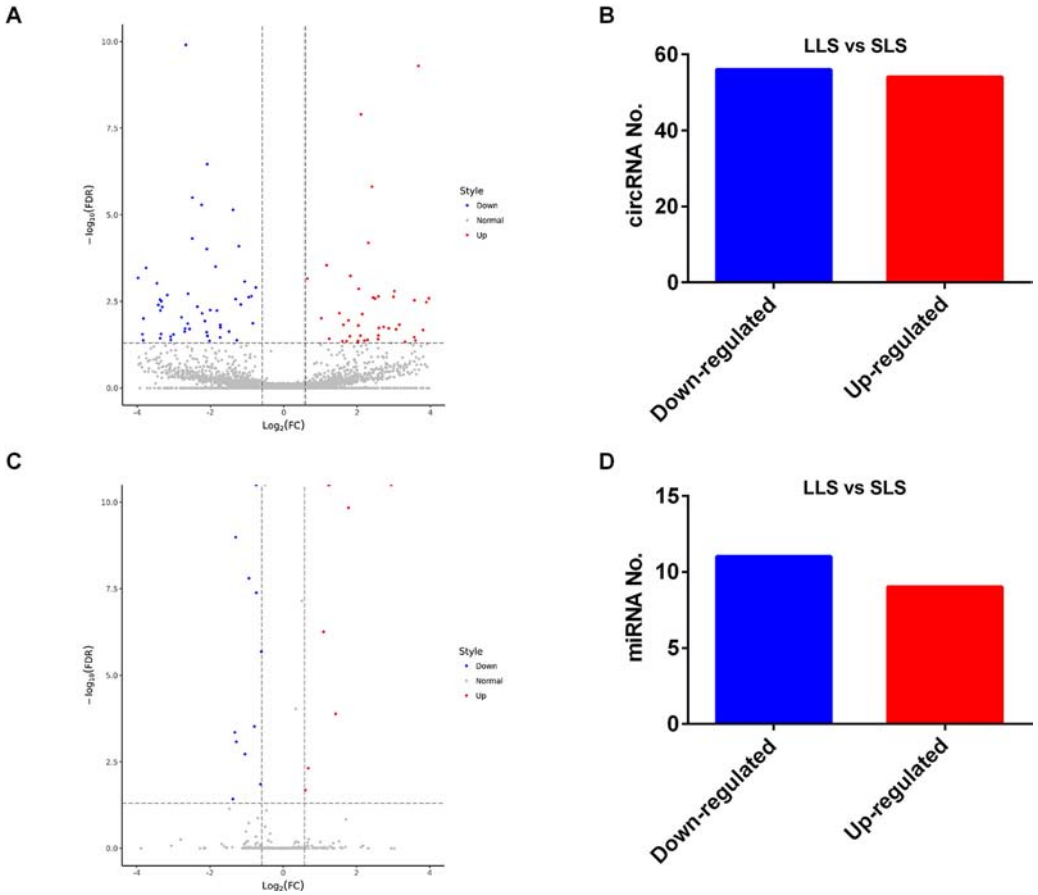


FIGURE 2 | Differentially expressed circRNAs and miRNAs between SLS and LLS pigs. **(A)** The volcano plot of differentially-expressed circRNAs in LLS and SLS ovaries. **(B)** The numbers of down- and up-regulated circRNAs in LLS ovaries compared with SLS. **(C)** The volcano plot of differentially-expressed miRNAs in LLS and SLS ovaries. **(D)** The numbers of down- and up-regulated miRNAs in LLS ovaries compared with SLS.

circ-*SNTB2*, circ-*TCP1*, and circ-*KMT2A* were significantly higher expressed in porcine ovaries with SLS, while circ-*CCDC85A* and circ-*CCAR1* were significantly higher expressed in porcine ovaries with LLS (**Figure 3**).

Regarding the differentially expressed miRNAs, the levels of miR-183 and miR-7857-3p were significantly lower in the SLS group while miR-497-5p were significantly lower expressed in pigs with LLS, which were consistent with high-throughput sequencing(**Figure 4**).

Identification of CircRNA–MiRNA Axis

Among the differentially expressed circRNAs and miRNAs detected above, RNAhybrid analysis revealed that miR-183 was predicted to interact with circ-*TCP1*, miR-497 with circ-*CCDC85A*. Besides, the expression of miR-183 was reversely correlated with circ-*TCP1*, and a similar tendency was observed between miR-497 and circ-*CCDC85A*. Meanwhile, dual-luciferase reporter assay has shown that miR-183 could directly bind to circ-*TCP1*, while the direct interaction between miR-497 and circ-*CCDC85A* was not detected in this assay (**Figure 5**).

Function Analysis of miR-183

TargetScan and MiRDB were used to predict the potential targets of miR-183, and the common genes presented by these two strategies were subjected to KEGG and GO analysis. KEGG showed that miR-183 might modulate DNA-templated and RNA PolII-mediated transcription (**Figure 6A**). GO enrichment assay indicated that miR-183 might be tightly related with PI3K-Akt signaling activity (**Figure 6B**).

DISCUSSION

The aim of the current study was to identify potential circRNAs related with swine fertility. Even RNA-seq only screened out a total of 1,291 exon-derived circRNAs that expressed in each ovary in our study; there were still several circRNAs that differentially expressed in ovaries with small and large litter size. To overcome the deficiency of great individual variation, the expressions of circRNAs of interest were further validated by RT-qPCR on larger-scale samples. Of note, significantly more circ-*TCP1* was detected in porcine ovaries with SLS, and we focus on circ-*TCP1* in the subsequent study.

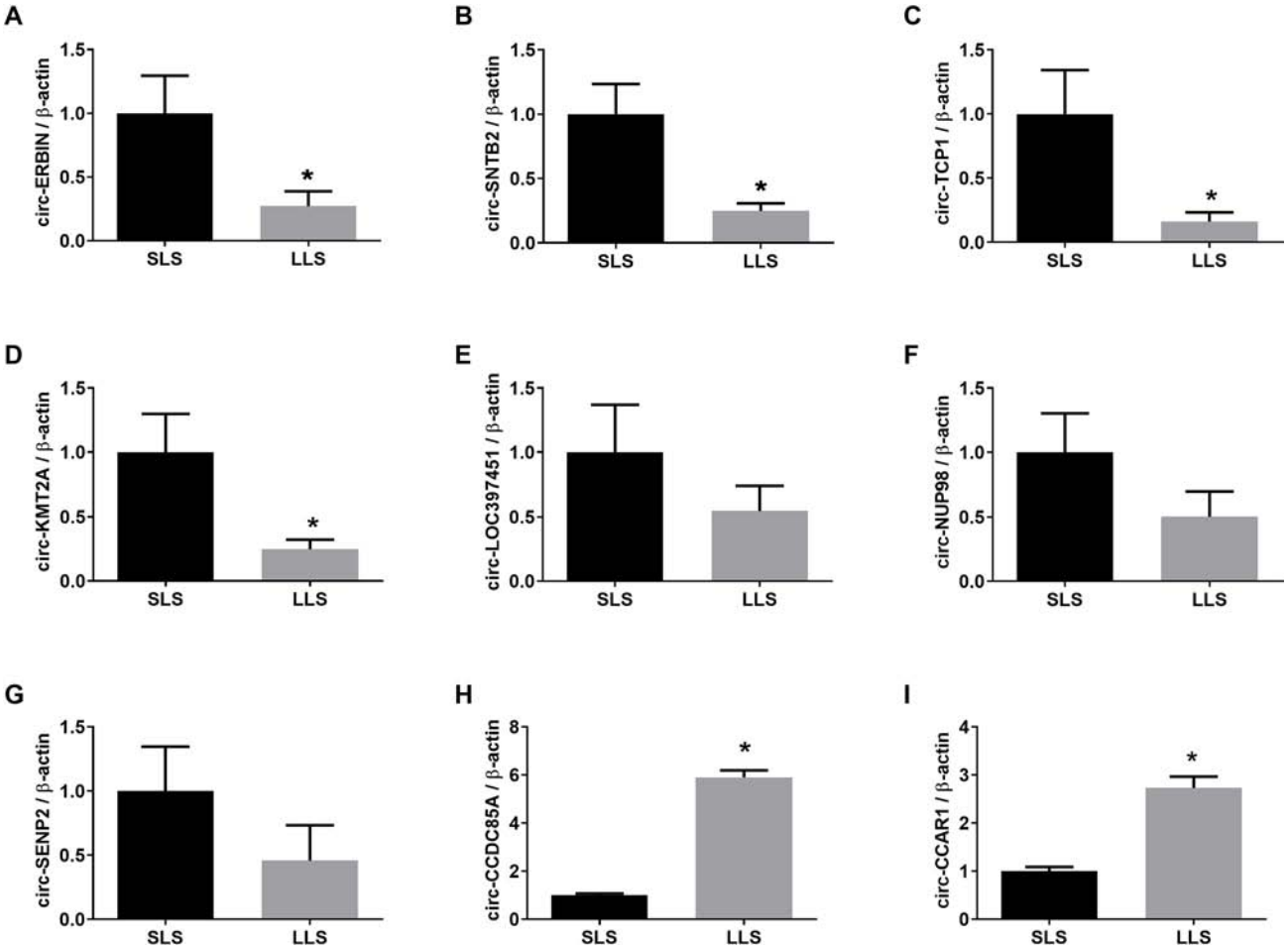


FIGURE 3 | Validation of differentially expressed circRNAs by qPCR. The expression of circ-ERBIN **(A)**, circ-SNTB2 **(B)**, circ-TCP1 **(C)**, circ-KMT2A **(D)**, circ-LOC397451 **(E)**, circ-NUP98 **(F)**, circ-SENP2 **(G)**, circ-CCDC85A **(H)** and circ-CCAR1 **(I)** in SLS and LLS ovaries. *presents $p < 0.05$.

One of the well-documented pathways of circRNAs is to competitively bind to functional miRNAs, known as competing endogenous RNAs (ceRNAs) (Li et al., 2018). Here, circ-*TCP1*, derived from exons 7 and 8 of porcine *TCP1* (T-complex protein 1 subunit alpha) gene, was significantly lower expressed in ovaries with LLS. *TCP1* gene encodes a molecular chaperone that is a member of the chaperonin containing TCP1 complex (CCT), also known as the TCP1 ring complex (TRiC), which folds various proteins, including actin and tubulin (Sternlicht et al., 1993). Currently, there are no reports about the role of *TCP1* in ovary. However, *CCT6A*, the zeta subunit of CCT, was shown to be expressed in chicken granulosa cells, indicating an important role in folic growth (Wei et al., 2013). Here, circ-*TCP1* was predicted to absorb miR-183 by online RNAhybrid software. Moreover, miR-183 presented a significantly reverse profile with circ-*TCP1*, and the interaction between circ-*TCP1* and miR-183 was further confirmed by luciferase activity assay. Collectively, our data indicated that the circ-*TCP1*–miR-183 axis might be involved in the biological processes related with litter size.

MiR-183 belongs to the highly conserved miR-183-96-182 cluster, which have been shown to be associated with female fertility (Zhang et al., 2019a). Members of miR-183-96-182 cluster are known to target the 3′-UTR of *FOXO1*, an important transcription factor for follicle-stimulating hormone responsive genes in ovarian granulosa cells of rodents (Herndon et al., 2016). *FOXO1* and miR-183-96-182 cluster have been also shown to be associated with bovine ovarian follicle development (Zielak-Steciwko and Evans 2016). MiR-183 is highly expressed in ovarian cancer cells (Wang et al., 2014; Chen et al., 2016) and down-regulation of miR-183 markedly represses cell proliferation and promotes apoptosis *via* targeting SMAD family member 4 (*Smad4*) (Zhou et al., 2019). In our study, KEGG and GO assay suggested that miR-183 might be associated with gene transcription, especially related with PI3K-Akt signaling. Genome-wide analysis

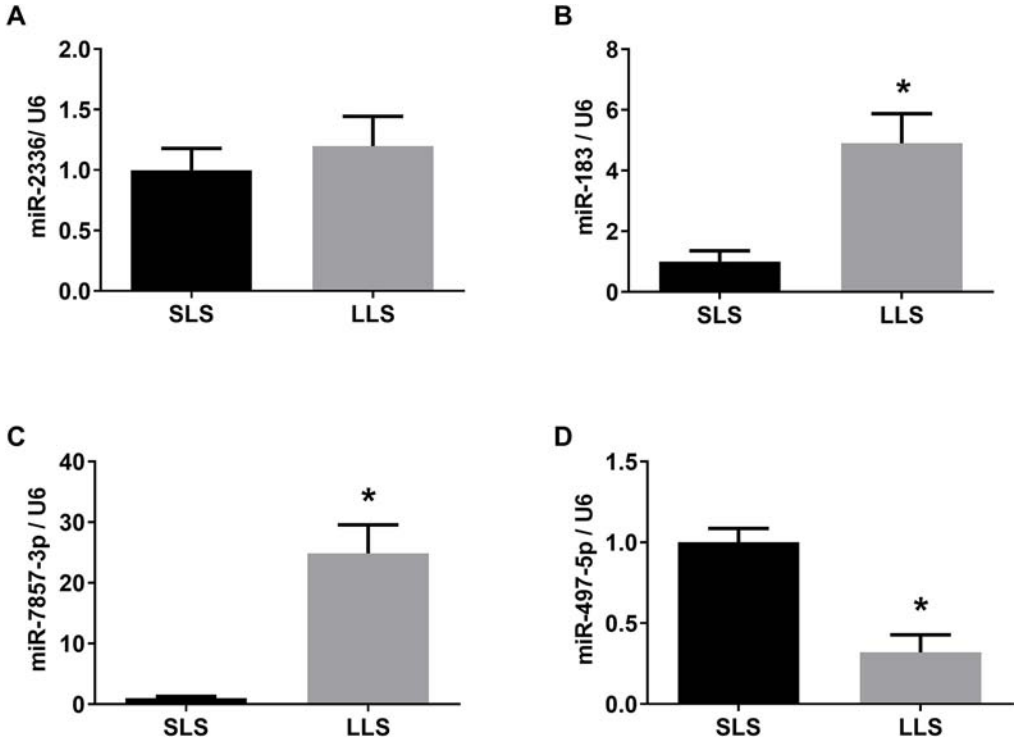


FIGURE 4 | Validation of differentially expressed miRNAs by qPCR. The expression of miR-2336 **(A)**, miR-183 **(B)**, miR-7857 **(C)** and miR-497-5p **(D)** in SLS and LLS ovaries. *presents $p < 0.05$.

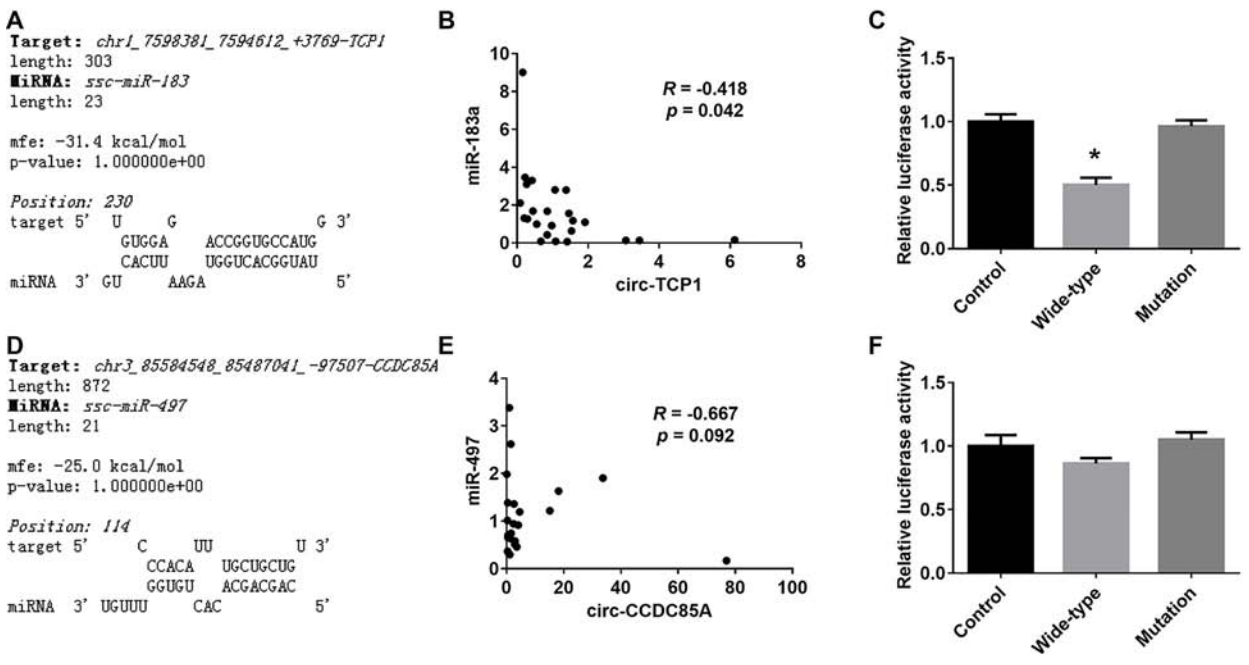


FIGURE 5 | Validation of the interaction between the differentially expressed miRNAs and circRNAs. **(A)** The putative binding site of miR-183 on circ-TCP1 predicted by RNAhybrid. **(B)** The correlation of the expression levels between miR-183 and circ-TCP1 in ovaries. **(C)** The relative luciferase activity of reporter constructs containing wild-type or mutant miR-183-binding sites from circ-TCP1 following cotransfection with control or miR-183 mimics. **(D)** The putative binding site of miR-497 on circ-CCDC85A predicted by RNAhybrid. **(E)** The correlation of the expression levels between miR-497 and circ-CCDC85A in ovaries. **(F)** The relative luciferase activity of reporter constructs containing wild-type or mutant miR-497-binding sites from circ-CCDC85A following cotransfection with control or miR-497 mimics. *presents $p < 0.05$.

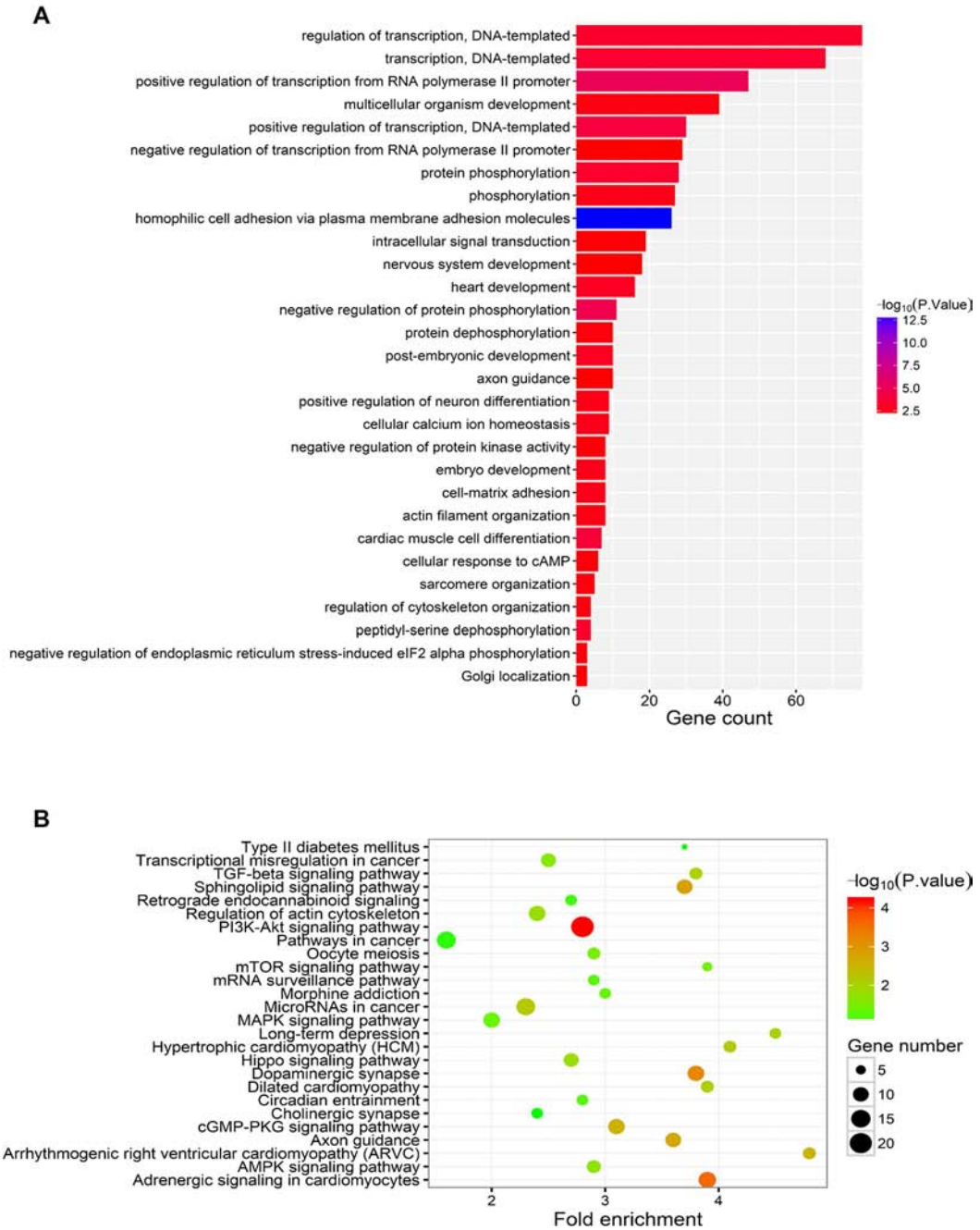


FIGURE 6 | Functional analysis of miR-183. **(A)** The KEGG pathway analysis of miR-183 target genes. **(B)** The GO enrichment analysis of miR-183 target genes.

revealed that genes related with PI3K-Akt activity were changed during different follicular stages in the ovaries of Duroc pigs (Liu et al., 2018), and PI3K-Akt activity was significantly inhibited when cell growth of porcine ovarian granulosa was impaired by extracellular stimuli (Wang et al., 2019; Zhang et al., 2019b). Altered PI3K-Akt signaling was also reported to contribute to impeded 17β-estradiol secretion in ovary cells (Wu et al., 2017). However, the molecular mechanism of miR-183-PI3K-Akt axis in ovary and their effects on litter size requires further investigation.

CONCLUSION

In our study, genome-wide identification of exon-derived circRNAs in porcine ovaries was performed by RNA-seq, and many of which were

differently expressed in ovaries with variant litter sizes. Furthermore, most exonic circRNAs harbored miRNA binding sites, and circ-*TCP1*-miR-183 axis might be associated with swine litter size.

DATA AVAILABILITY STATEMENT

The data generated in this study has been uploaded to NCBI and can be found under accession number: GSE136592

ETHICS STATEMENT

This study was approved by Animal Care and Use Committee in Northwest A&F University.

AUTHOR CONTRIBUTIONS

GX conducted the study and drafted the manuscript. HZ and XL assisted t in ovaries sampling and data analysis. JH, GY gave critical comments about experiment design and manuscript drafting. SS supervised the experiment.

FUNDING

This work was supported by the National Key Technology R and D Program of China (2015BAD03B01-10).

REFERENCES

Cai, H., Li, Y., Li, H., Niringiyumukiza, J. D., Zhang, M., et al. (2018). Identification and characterization of human ovary-derived circular RNAs and their potential roles in ovarian aging. *Aging* 10, 2511–2534. doi: 10.18632/aging.101565

Chen, H., Zhang, L., Zhang, L., Du, J., Wang, H., et al. (2016). MicroRNA-183 correlates cancer prognosis, regulates cancer proliferation and bufalin sensitivity in epithelial ovarian caner. *Am. J. Transl. Res.* 8, 1748–1755.

Herndon, M. K., Law, N. C., Donaubauer, E. M., Kyriss, B., and Hunzicker-Dunn, M. (2016). Forkhead box O member FOXO1 regulates the majority of follicle-stimulating hormone responsive genes in ovarian granulosa cells. *Mol. Cell. Endocrinol.* 434, 116–126. doi: 10.1016/j.mce.2016.06.020

Huang, L., Yin, Z. J., Feng, Y. F., Zhang, X. D., Wu, T., et al. (2016). Identification and differential expression of microRNAs in the ovaries of pigs (Sus scrofa) with high and low litter sizes. *Animal Genet.* 47, 543–551. doi: 10.1111/age.12452

Kemp, B., Da Silva, C. L. A., and Soede, N. M. (2018). Recent advances in pig reproduction: Focus on impact of genetic selection for female fertility. *Reprod. Domestic Animals* 53, 28–36. doi: 10.1111/rda.13264

Li, X., Yang, L., and Chen, L.-L. (2018). The biogenesis, functions, and challenges of circular RNAs. *Mol. Cell* 71, 428–442. doi: 10.1016/j.molcel.2018.06.034

Liang, G., Yang, Y., Niu, G., Tang, Z., and Li, K. (2017). Genome-wide profiling of Sus scrofa circular RNAs across nine organs and three developmental stages. *DNA Res.* 24, 523–535. doi: 10.1093/dnares/dsx022

Liu, Y., Li, M., Bo, X., Li, T., Ma, L., et al. (2018). Systematic analysis of long non-coding RNAs and mRNAs in the ovaries of Duroc pigs during different follicular stages using RNA sequencing. *Int. J. Mol. Sci.* 19, E1722. doi: 10.3390/ijms19061722

Quan, G., and Li, J. (2018). Circular RNAs: biogenesis, expression and their potential roles in reproduction. *J. Ovarian Res.* 11, 9–9. doi: 10.1186/s13048-018-0381-4

Sternlicht, H., Farr, G. W., Sternlicht, M. L., Driscoll, J. K., Willison, K., et al. (1993). The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin *in vivo*. *Proc. Natl Acad. Sci. U. S. A.* 90, 9422–9426. doi: 10.1073/pnas.90.20.9422

Tang, L.-T., Ran, X.-Q., Mao, N., Zhang, F.-P., Niu, X., et al. (2018). Analysis of alternative splicing events by RNA sequencing in the ovaries of Xiang pig at estrous and diestrous. *Theriogenology* 119, 60–68. doi: 10.1016/j.theriogenology.2018.06.022

Wang, L., Zhu, M.-J., Ren, A.-M., Wu, H.-F., Han, W.-M., et al. (2014). A ten-microRNA signature identified from a genome-wide microRNA expression profiling in human epithelial ovarian cancer. *PloS One* 9, e96472–e96472. doi: 10.1371/journal.pone.0096472

Wang, W., Luo, S., Ma, J., Shen, W., and Yin, S. (2019). Cytotoxicity and DNA damage caused from Diazinon exposure by inhibiting the PI3K-AKT pathway in porcine ovarian granulosa cells. *J. Agric. Food Chem.* 67, 19–31. doi: 10.1021/acs.jafc.8b05194

Wei, Q., Zhu, G., Cui, X., Kang, L., Cao, D., et al. (2013). Expression of CCT6A mRNA in chicken granulosa cells is regulated by progesterone. *Gen. Comp. Endocrinol.* 189, 15–23. doi: 10.1016/j.ygcen.2013.04.019

Wu, C., Yan, D., Lu, D., Han, T., and Zhao, B. (2017). Alteration of the PI3K/Akt signaling pathway by swainsonine affects 17β-Estradiol secretion in ovary cells. *Theriogenology* 103, 123–129. doi: 10.1016/j.theriogenology.2017.07.033

Zak, L. J., Gaustad, A. H., Bolarin, A., Broekhuijse, M. L. W. J., Walling, G. A., et al. (2017). Genetic control of complex traits, with a focus on reproduction in pigs. *Mol. Reprod. Dev.* 84, 1004–1011. doi: 10.1002/mrd.22875

Zhang, J., Xu, Y., Liu, H., and Pan, Z. (2019a). MicroRNAs in ovarian follicular atresia and granulosa cell apoptosis. *Reprod. Biol. Endocrinol.: RB&E* 17, 9–9. doi: 10.1186/s12958-018-0450-y

Zhang, T., Sun, X., Li, L., Ma, J., Zhang, R., et al. (2019b). Ochratoxin A exposure impairs porcine granulosa cell growth via the PI3K/AKT signaling pathway. *J. Agric. Food Chem.* 67, 2679–2690. doi: 10.1021/acs.jafc.8b06361

Zhang, X., Huang, L., Wu, T., Feng, Y., Ding, Y., et al. (2015). Transcriptomic analysis of ovaries from pigs with high and low litter size. *PLoS One* 10, e0139514. doi: 10.1371/journal.pone.0139514

Zhou, J., Zhang, C., Zhou, B., and Jiang, D. (2019). miR-183 modulated cell proliferation and apoptosis in ovarian cancer through the TGF-β/Smad4 signaling pathway. *Int. J. Mol. Med.* 43, 1734–1746. doi: 10.3892/ijmm.2019.4082

Zielak-Steciwko, A. E., and Evans, A. C. O. (2016). Genomic portrait of ovarian follicle growth regulation in cattle. *Reprod. Biol.* 16, 197–202. doi: 10.1016/j.repbio.2016.07.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Xu, Zhang, Li, Hu, Yang and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 November 2019
doi: 10.3389/fgene.2019.01130

Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle

*Wietje Nolte[1], Rosemarie Weikard[1], Ronald M. Brunner[1], Elke Albrecht[2], Harald M. Hammon[3], Antonio Reverter[4] and Christa Kühn[1,5]**

[1] Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, [2] Institute of Muscle Biology and Growth, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, [3] Institute of Nutritional Physiology "Oskar Kellner," Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, [4] Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, Queensland Bioscience Precinct, St Lucia, QLD, Australia, [5] Faculty of Agricultural and Environmental Sciences, University Rostock, Rostock, Germany

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
James Reecy,
Iowa State University,
United States
Kieran G. Meade,
The Irish Agriculture and Food Development Authority, Ireland

***Correspondence:**
Christa Kühn
kuehn@fbn-dummerstorf.de

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 25 June 2019
Accepted: 17 October 2019
Published: 22 November 2019

Citation:
Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A and Kühn C (2019) Biological Network Approach for the Identification of Regulatory Long Non-Coding RNAs Associated With Metabolic Efficiency in Cattle. Front. Genet. 10:1130. doi: 10.3389/fgene.2019.01130

Background: Genomic regions associated with divergent livestock feed efficiency have been found predominantly outside protein coding sequences. Long non-coding RNAs (lncRNA) can modulate chromatin accessibility, gene expression and act as important metabolic regulators in mammals. By integrating phenotypic, transcriptomic, and metabolomic data with quantitative trait locus data in prioritizing co-expression network analyses, we aimed to identify and functionally characterize lncRNAs with a potential key regulatory role in metabolic efficiency in cattle.

Materials and Methods: Crossbred animals (n = 48) of a Charolais x Holstein F_2-population were allocated to groups of high or low metabolic efficiency based on residual feed intake in bulls, energy corrected milk in cows and intramuscular fat content in both genders. Tissue samples from jejunum, liver, skeletal muscle and rumen were subjected to global transcriptomic analysis via stranded total RNA sequencing (RNAseq) and blood plasma samples were used for profiling of 640 metabolites. To identify lncRNAs within the indicated tissues, a project-specific transcriptome annotation was established. Subsequently, novel transcripts were categorized for potential lncRNA status, yielding a total of 7,646 predicted lncRNA transcripts belonging to 3,287 loci. A regulatory impact factor approach highlighted 92, 55, 35, and 73 lncRNAs in jejunum, liver, muscle, and rumen, respectively. Their ensuing high regulatory impact factor scores indicated a potential regulatory key function in a gene set comprising loci displaying differential expression, tissue specificity and loci overlapping with quantitative trait locus regions for residual feed intake or milk production. These were subjected to a partial correlation and information theory analysis with the prioritized gene set.

Results and Conclusions: Independent, significant and group-specific correlations ($|r| > 0.8$) were used to build a network for the high and the low metabolic efficiency group resulting in 1,522 and 1,732 nodes, respectively. Eight lncRNAs displayed a particularly high connectivity (>100 nodes). Metabolites and genes from the partial correlation and information theory networks, which each correlated significantly with the respective lncRNA, were included in an enrichment analysis indicating distinct affected pathways

for the eight lncRNAs. LncRNAs associated with metabolic efficiency were classified to be functionally involved in hepatic amino acid metabolism and protein synthesis and in calcium signaling and neuronal nitric oxide synthase signaling in skeletal muscle cells.

Keywords: *Bos taurus*, metabolic efficiency, co-expression network analysis, long non-coding RNA, Functional Annotation of Animal Genomes

INTRODUCTION

In recent years the focus of livestock production and farming has shifted in developed countries towards a stronger emphasis on resource efficiency and sustainability (Thornton, 2010). In cattle, energy metabolism, nutrient conversion and efficient use of primary resources are of increasing economic and ecological importance to breeders and consumers. Genomic selection and the use of biomarkers greatly facilitate the improvement of complex phenotypes, e.g. feed efficiency, which remain cost- and time-consuming to measure (Kenny et al., 2018).

Some pivotal gene mutations are known in major livestock production traits, e.g. a meta-analysis on stature in cattle identified *PLAG1* as a major regulator and pointed towards putative causal mutations (Bouwman et al., 2018). In pigs, the scavenger receptor cysteine-rich domain 5 in gene *CD163*, when not being translated, led to resistance to porcine reproductive and respiratory syndrome virus 1 infection (Burkard et al., 2018). Pigs that did not express the receptor protein were susceptible to the infection. For the region between *LCORL* and *NCAPG*, which has been associated with growth or feed efficiency in a number of species (cattle, horse, human), multiple mappings have narrowed down the region of interest but the causal mutation remains unknown (Widmann et al., 2015; Bouwman et al., 2018). A large part of the variation in traits like feed efficiency, growth and carcass traits remains still unexplained (Hardie et al., 2017; Medeiros de Oliveira Silva et al., 2017; Seabury et al., 2017) and genome-wide association studies repeatedly pointed towards quantitative trait loci (QTL) outside protein-coding genes (Ibeagha-Awemu et al., 2016; Seabury et al., 2017; Higgins et al., 2018).

Due to their gene expression regulatory potential, long non-coding RNAs (lncRNAs) have emerged as potential key regulators for diverse biological processes, such as X-chromosomal inactivation and dosage compensation (Brown et al., 1992; Clemson et al., 1996), vernalization/ flowering in plants (Csorba et al., 2014), as well as human cancer biology as reviewed by Serviss et al. (2014).

Recently, lncRNAs have been suggested as therapeutic targets for diabetes and other metabolic diseases because of their involvement in lipid metabolism, adipogenesis and fat deposition (Chen et al., 2018a; Liu et al., 2018; Zeng et al., 2018). In mammals, lncRNAs were further identified as key regulators of energy metabolism and lipogenesis (Yang et al., 2016). In adipocytes, these genomic elements also play an integral part in the insulin-signaling pathway (Degirmenci et al., 2019). A central regulatory role of lncRNAs was furthermore observed in skeletal muscle in myogenesis and muscle cell differentiation: *SYISL* has been shown to regulate myoblast proliferation and fusion and acts in an inhibitory way in myogenic differentiation (Jin et al., 2018), *Irm* enhances myogenic differentiation during myogenesis through the binding to *MEF2D* (Sui et al., 2019), and *lnc-mg* overexpression has directly been linked to muscle hypertrophy in mice, whereas a knock-out led to dystrophy (Zhu et al., 2017). It is likely that lncRNAs contribute significantly to economically important production traits and divergent phenotypes in livestock as well. Since they show little sequence conservation across species and their expression appears to be mainly species specific and spatiotemporal (Ulitsky et al., 2011; Ulitsky and Bartel, 2013), knowledge transfer remains a challenging issue. The identification and functional characterization of lncRNAs needs to be performed for each species, and this fits into one of the major goals of the consortium for the Functional Annotation of Animal Genomes (FAANG, https://www.animalgenome.org/community/FAANG/) that strives to identify and annotate functionally relevant elements in livestock genomes.

Another key feature of lncRNAs is their low expression level compared to protein-coding genes (Derrien et al., 2012), which makes their detection challenging. From transcription factors it is known, that little changes in abundance can however have tremendous consequences if these have high regulatory potential in terms of gene expression (Vaquerizas et al., 2009) and we postulated an analogous phenomenon for lncRNAs. For instance, the knockout of the lowly expressed lncRNA *ßlinc* in mice impaired the correct formation of pancreatic islets and severely changed the glucose homeostasis in adult animals (Arnes et al., 2016). A low and tightly regulated gene expression has implications for differential expression (DE) analyses, because little changes in expression are often not recognized as significant due to lack of power in standard experimental designs. Therefore, other approaches are necessary when aiming to identify and functionally annotate key regulatory lncRNAs. A tested and proven method in the screening for critical transcription factors from gene expression data, which are typically low in abundance but have high regulatory power as reviewed by Vaquerizas et al. (2009), is network co-expression analysis that incorporates the regulatory impact factor (RIF) metrics and a partial correlation and information theory (PCIT) (Reverter et al., 2010; Perez-Montarelo et al., 2012). This approach has previously also led to the identification of regulatory elements associated with puberty (Canovas et al., 2014; Nguyen et al., 2018) and feed efficiency in cattle (Alexandre et al., 2019). We assumed that this rational network approach could also be used as a hypothetical generation tool for the systematic detection of lncRNAs with important regulatory potential.

In this study, we took advantage of a unique F_2 cross-population of meat and dairy cattle breeds (Charolais x Holstein) (Kühn et al., 2002) that has been deeply phenotyped and genotyped.

Earlier studies have shown that in this cross population a gene variant of the *NCAPG* gene is associated with fetal and pubertal growth (Eberlein et al., 2009; Weikard et al., 2010). By integrating quantitative metabolite data with genotype information, this *NCAPG* genotype was found to be associated with plasma arginine levels (Weikard et al., 2010). A systems biology approach, which combined metabolome data, growth-associated phenotypic and genetic information, revealed a functional gene interaction network characterizing the intensive growth phase at the beginning of the pubertal growth interval (Widmann et al., 2013). Potential interaction partners of the *NCAPG* gene were predicted and the functional role of the *NCAPG* gene as a growth regulator linked to the arginine NO metabolism was concluded. A combined phenotype–metabolome–genome analysis was also used to identify genetic switches of associated molecular signaling pathways linked to variance in efficiency of feed conversion (Widmann et al., 2015).

This current study on the regulatory role of lncRNAs for metabolic efficiency was aimed to contribute to a more detailed elucidation of the molecular background of this complex physiological trait and help to characterize divergent metabolic types with respect to nutrient partitioning. Therefore, phenotypic information, transcriptomic data from four metabolically relevant tissues and QTL information were used to establish a prioritized gene set that was submitted to the combinational RIF metrics and subsequently to the PCIT algorithm for co-expression network creation. The integration of metabolomic profiles through correlation with transcriptomic data added valuable information for the interpretation of biological functions.

MATERIALS AND METHODS

Design of the Study

For this study, we made use of 48 animals (24 bulls, 24 cows) of a F_2-population [SEGFAM (Kühn et al., 2002)] from a Charolais × Holstein cross. The cross population was bred at the Leibniz Institute for Farm Animal Biology in Dummerstorf (Germany) and kept under standardized housing and feeding conditions as previously described (Eberlein et al., 2009; Weikard et al., 2010; Widmann et al., 2011). Males were slaughtered at 18 months of age and females were slaughtered after their second parity at 30 days postpartum. Based on residual feed intake (RFI) in bulls and energy corrected milk yield (ECM_w) in cows as well as intramuscular fat content (IMF) of *M. longissimus dorsi* in both genders, animals were assigned to either of the two groups: high or low metabolic efficiency (**Table 1**). In this study we defined high metabolic efficiency in cattle as the preference to accrete or secrete protein while receiving the same diet as their inefficient conspecifics, which were characterized by a clear tendency to accrete fat instead of protein. In European production systems, those animals are most sustainable and economically efficient producers, which build up protein mass (muscle) with little fat content or, in case of females, secrete high amounts of milk.

Cows were categorized as highly efficient if their milk yield within the 7 days prior to slaughter was above 140 kg energy correct milk (ECM_w) and the carcass fat content (CFC) was less than the average CFC of all cows plus one standard deviation. In contrast, cows were classified as lowly efficient if their milk yield within the last week was between 14 and 40 kg ECM_w and the CFC was above the average CFC of all cows minus one standard deviation. For all cows, the calving interval had to be less than 540 days, the maximum age was 1,510 days and they had to be free of pathological findings with metabolic implications noted after slaughter. Cows that were categorized as highly efficient (high ECM_w) on average had a lower CFC (mean 17.1%, SD 2.7%) and lowly efficient cows (low ECM_w) had a higher CFC (mean 25.9%, SD 3.6%) than the mean of the population (21.8%, SD 5.3%, n = 242). In addition, highly efficient cows had a lower IMF (mean 4.16%, SD 1.60%) and the lowly efficient cows had a higher IMF (mean 6.46%, SD 2.53%) than the mean of the population (5.21%, SD 2.21%, n = 242).

The individual milk volume yield per cow was measured on a daily basis and the milk composition was determined once per week. The trait included in cow selection for this study corresponded to the weekly ECM determined for the 7 days before slaughter (ECM_w). The formula presented by Kirchgeßner (1997) was modified accordingly for the one week interval (F% = milk fat percentage, P% = milk protein percentage):

$$ECM_w = \frac{0.37\ F\% + 0.21\ P\% + 0.95}{3.1} \times MY - 7d$$

cows, the ECM_w was used as a substitute feature for feed efficiency, because the facilities did not allow for RFI measurement in cows during the time of the experiment.

TABLE 1 | Sample characteristics.

Metabolic efficiency group	Number of animals	Sex	RFI[1] in last month of life (bulls)	ECM^2_w (cows)	IMF[3] (both sexes)	CFC[4] (both sexes)
			μ[5] (SD[6])	μ (SD)	μ (SD)	μ (SD)
High	25	12 males 13 females	-21.30 (4.44)	190.87 (22.02)	3.46 (1.30)	15.93 (3.16)
Low	23	12 males 11 females	20.83 (4.41)	30.97 (9.18)	5.51 (2.34)	22.93 (4.88)

[1]RFI, residual feed intake; [2]ECM_w, energy corrected milk 7 days before slaughter; [3]IMF, intramuscular fat content (given in percent, measured in M. longissimus dorsi); [4]CFC, carcass fat content; [5]μ, mean; [6]SD, standard deviation.

For bulls, the decisive factor for animal selection was RFI calculated for the last month prior to slaughter. The RFI equals the animals' energy intake while considering the average daily gain and metabolic mid-weight (average body weight of months of life 17 to 18 raised to the power of 0.75) (Archer et al., 1997).

Bulls with a low RFI (at least 1 standard deviation below average) were assigned to the high metabolic efficiency group and bulls with a high RFI (at least one standard deviation above average) were assigned to the low metabolic efficiency group. In their last month of life, all bulls had to have a positive daily weight gain and no less than the population average minus one standard deviation. Bulls that were categorized as highly efficient (negative RFI) on average had a lower CFC (mean 14.2%, SD 3.0%) and lowly efficient bulls (positive RFI) had a higher CFC (mean 20.2%, SD 4.4%) than the population mean (mean 16.5%, SD 4.0%, n = 246). Analogously to cows, highly efficient bulls had a lower IMF (mean 1.71%, SD 1.00%) and the lowly efficient bulls had a higher IMF (mean 4.64%, SD 1.84%) than the population mean (mean 3.67%, SD 1.76%, n = 246).

Plasma Metabolic Profiles

Blood samples were collected from all individuals (n = 48) at slaughter. Plasma samples were sent to Metabolon Inc. (Durham/NC, USA) for the establishment of holistic metabolite profiles that included 640 biochemical compounds and molecules. Metabolites with more than five animals with missing data were excluded. After this filtering step, 490 metabolites remained and missing values were imputed with the minimum measurement, assuming that missing values were due to concentrations below the detection limit. Values were then scaled without centering for each metabolite in R (Core Team 2018) with the scale-function.

All experimental procedures were carried out according to the German animal care guidelines and were approved and supervised by the relevant authorities of the State Mecklenburg-Vorpommern, Germany (State Office for Agriculture, Food Safety and Fishery; LALLF M-V/TSD/7221.3-2.1-010/03).

Sampling, RNA Isolation, Library Preparation, and Sequencing

Tissue samples were collected from jejunum mucosa, liver (*Lobus caudatus*), skeletal muscle *(M. longissimus dorsi*), and rumen (*Saccus ventralis*, papillary base) directly after slaughtering and dissection, shock frozen in liquid nitrogen and subsequently stored at -80°C.

For RNA extraction from muscle and rumen, frozen samples (100 mg) were treated with 1 ml TRIzol reagent (Invitrogen, Darmstadt, Germany) and subjected to the Precellys-24 homogenizer (5,500 rpm, 2 × 15 s, lysing kit containing 1.4 mm ceramic beads). For RNA extraction from liver and jejunum, frozen tissue samples were grinded in liquid nitrogen and 30 mg were used for further purification steps. No TRIzol was used for liver and jejunum samples. All samples were then subjected to an on-column-purification step with the NucleoSpin RNA II kit (Macherey & Nagel, Düren, Germany) including a DNase digestion to remove genomic DNA. In addition, the RNA was tested for remaining traces of DNA contamination and, in case of remaining DNA residues, further cleansed according to Weikard et al. (2012).

The RNA concentration and integrity were measured with a Qubit Fluorometer (Invitrogen, Germany) and a 2100 Bioanalyzer Instrument (Agilent Technologies, Germany). Stranded, ribodepleted and indexed libraries were prepared from 1 µg total RNA using the TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold Kit (Illumina, San Diego, USA) and subjected to paired-end sequencing (2 × 100 bp) in a multiplexed design on a HiSeq 2500 Sequencing System (Illumina).

Alignment and Assembly

After quality control of raw sequencing reads with FastQC (Andrew, 2010), adapter and quality trimming were performed with Cutadapt v. 1.16 (Martin, 2011) and Quality Trim v. 1.6.0 (Robinson, 2015), respectively. In Quality Trim the start of sequences was also trimmed (option -s) and the maximum number of N bases was set to 3, while the minimum base quality was set to 15. Reads were then mapped in a guided alignment with HISAT2 v.2.1.0 (Kim et al., 2015) to the bovine reference genome UMD.3.1 [Ensembl annotation release 92 (Frankish et al., 2017)]. After sorting and indexing of BAM files with samtools v.1.6 (Li et al., 2009), samples were individually assembled with Stringtie v.1.3.4d (Pertea et al., 2015) based on the reference genome and annotation used for alignment. Using the individually assembled samples (n = 204) from all four tissues and the bovine reference genome, we built a new merged annotation in Stringtie across tissues, while specifying for minimal transcript coverage across samples of 15 read alignments per exonic base. In addition to the 192 samples (48 animals, four tissues) included in the subsequent steps for DE and network analyses, we also took benefit from rumen, liver and muscle samples of further four individuals from the same experimental herd. These samples were subjected to exactly the same processing steps as the 192. The new merged annotation was used for fragment counting with featureCounts (subread v.1.6.1) (Liao et al., 2014), while allowing for fractional counting and specifying for reverse strandedness.

Long Non-Coding RNA Prediction and Fragment Counting

LncRNAs were identified *in-situ* with FEELnc (Wucher et al., 2017), a bioinformatics tool for lncRNA prediction and annotation, using the merged transcript annotation and the bovine reference genome and annotation UMD3.1 release 92. FEELnc excludes transcripts annotated as protein coding and subsequently keeps transcripts with a minimum length of 200 nt and at least two exons and only monoexonic transcripts with antisense localization. Other monoexonic transcripts were excluded to reduce the number of false positives, which might arise from the mapping of repetitive sequences (Wucher et al., 2017), DNA contamination (Haerty and Ponting, 2015) and in general transcriptional noise (Kern et al., 2018). For those transcripts matching the requirements, the coding potential of remaining transcripts was determined in shuffling mode.

Fragment Count Normalization

For further pipeline steps, except for the DE analysis, fragments per kilobase million (FPKM) were calculated from the featureCounts derived fragment counts. Genes were filtered for a minimal average expression value of 0.2 FPKM in at least one of the four tissues and ribosomal and spliceosomal RNA genes were excluded (Metazoan signal recognition particle RNA, U6 spliceosomal RNA, small nucleolar RNA U6-53). For further analyses of FPKM values performed in this study, a log2-scale of the data was used (for log transformation a pseudo-count of 0.001 was added).

Prioritized Gene List

Gene co-expression networks are a useful tool when trying to deduce the potential biological function of genes, novel loci and non-coding elements (van Dam et al., 2017), assuming the guilt-by-association principle. In order to create meaningful networks that have a targeted focus on our phenotype (metabolic efficiency), we created a set of prioritized genes where genes had to belong to at least one of these four categories: differentially expressed (DE) genes in at least one of the four investigated tissues, tissue-specific (TS) genes, genes harboring a QTL for milk production or RFI (QTL) according to the literature, and predicted lncRNAs. Small nucleolar RNAs (snoRNAs), ribosomal RNAs, spliceosomal RNAs, and Y-RNAs were excluded from the set.

Differential Expression Analysis

A DE analysis for the high and low metabolic efficiency group was performed within tissues and across sexes in R with the package DEseq2 (Love et al., 2014). Fragment counts from featureCounts were used as input and normalization was performed within DEseq2. To exclude very lowly expressed transcripts within a tissue, the minimal fragment count threshold was set to at least 10 fragments for 10 out of 48 individuals. Ribosomal genes were excluded from the analysis and year of slaughter and sex were used as factors in the model. The significance threshold was set to $q < 0.05$ [Benjamini–Hochberg (BH) test].

Tissue Enriched Genes

The expression (log2-transformed FPKM) of a gene was defined as enriched in a particular tissue, if the abundance in the other three tissues was less than half the average across all tissues and above the average plus one standard deviation in the tissue at hand. Throughout the further course of this study, we refer to these genes as TS.

Genes Harboring a Quantitative Trait Locus

We extracted QTL for milk production traits (MY) and RFI in cattle from the Animal QTL database (Park et al., 2018) and then screened our dataset in Ensembl Biomart (http://asia.ensembl.org/biomart/martview, accession date 28 March 2019) for genes that overlapped with these QTL regions. A physical overlap of the QTL and the gene is needed for a gene hit, while close neighborhood is not sufficient.

Regulatory Impact Factor Analysis

The RIF (Reverter et al., 2010) analysis makes use of two alternative metrics (RIF1 and RIF2) that attribute scores to potential key regulators. The strength of the score depends on the change in correlation between the regulator and its target in two groups or treatments, the level of DE of the target gene, and the general expression level of the target gene. We conducted RIF analyses within tissues and across metabolic efficiency groups to assess the regulatory capacity of lncRNAs in a set of prioritized genes (lncRNA, DE, TS, QTL harboring). Therefore, RIF metrics were calculated within each tissue for a prioritized gene set (including log2(FPKM) data) that comprised genes which were DE or TS in that tissue, harbored a QTL or were characterized as a lncRNA. Naturally, some of the QTL-genes might have zero expression in one or more of the tissues. To prevent erroneously high RIF scores stemming from low variation in gene expression, an additional filter for expression level was applied (on top of minimal average expression of 0.2 FPKM in at least one tissue). Only genes with abundance above tissue average were kept for the RIF analysis.

A high RIF1 score was assigned to lncRNAs that were consistently co-expressed with abundant target genes in both metabolic efficiency groups. A high RIF2 score was attributed to lncRNAs that displayed the most altered ability to predict the abundance of target genes between groups, meaning that a lncRNA exhibited strong correlation to a target on one condition but none or a reverse correlation in the other. RIF scores were standardized with a z-score. Key regulators (lncRNA) were considered of significant importance and were included in further analyses if they had an absolute RIF1 or RIF2 z-score of ≥ 1.96, meaning that these lncRNAs and their scores were outside the 95% confidence interval, corresponding to a significance level of $p = 0.05$ in a t-test.

Partial Correlation and Information Theory

The PCIT (Reverter and Chan, 2008) tests for significant pairwise correlations between two elements while accounting for all possible three-way combinations in the dataset that include either of the pair elements. Importantly, the PCIT recognizes independent, significant correlations regardless of the strength of correlation. Within the high and low metabolic efficiency groups, the PCIT approach across all tissues was used to investigate for independent correlations of lncRNAs that had significant RIF scores with DE genes, TS genes, and QTL harboring genes.

Results were filtered for significant correlations (minimal correlation strength $|r| > 0.8$) between a lncRNA and another gene that were exclusive for the high or low metabolic efficiency group, meaning that the correlation was significant in one group but not in the other. The visualization was realized in Cytoscape 3.6.1 (Shannon et al., 2003).

Characterization of Key Regulatory Long Non-Coding RNAs

Blast Search Against New Bovine Assembly

Highly connected lncRNAs with more than 100 directly linked nodes (genes) were selected from each network for further scrutiny. Since the prediction of lncRNAs was based on a merged annotation, which was reference guided by UMD3.1, Ensembl release 92, we wanted to investigate the sequence homology

and annotation status of key lncRNAs in the new bovine assembly ARS1.2 annotated in Ensembl release 95. The lncRNA sequences were blasted online with the blastn suite using the MegaBlast algorithm, specifying for high sequence similarity and otherwise default parameters (Altschul et al., 1990) (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed Mai 2019) against the new bovine assembly (ARS-UCD1.2, https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2; GenBank accession NKLS00000000.2; https://www.ensembl.org/Bos_taurus/Info/Index). We considered blast hits to indicate high homology if the sequence identity was at least 98% in a region covering at least 200 nucleotides.

Pathway Enrichment Analysis

To assess the possible biological function of high connectivity lncRNAs, we performed a pathway enrichment analysis based on genes identified as correlated ($|r| > 0.8$) in the PCIT analyses and also including blood plasma metabolites that were significantly ($p \leq 0.05$) correlated with the high connectivity lncRNAs. To this end, a pairwise Pearson correlation analysis between blood-plasma metabolites and lncRNA expression in the tissue, where the lncRNA was most abundant, was performed in R with the function rcorr of the Hmisc package (Harrell and Frank, 2019). The list of significantly correlated metabolites ($p \leq 0.05$) and genes (adjacent network nodes with $|r| > 0.8$) were analysed using the Ingenuity Pathway Analysis (IPA: QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) (Kramer et al., 2014). The workflow from group formation and tissue sampling up to the functional characterization of key lncRNAs is visualized for better comprehensibility and clarity in **Figure 1**.

RESULTS

RNA Preparation, Sequencing, Alignment, and Mapping

The average RNA integrity (RIN) across the four tissues was 8.22 ± 0.81 (**Table 2**). After quality trimming the average RNA sequencing depth was at 48 million read pairs per sample. A total of 9 out of 192 samples reached less than a 40 million read pair coverage. The alignment of reads with HISAT2 to the bovine reference genome UMD.3.1 (Ensembl release 92) resulted in an average alignment rate of 92.98 ± 9.50%. Compared with the other tissues, rumen scored a distinctly lower rate (78.00 ± 7.75%). The average mapping rate across all samples to the customized annotation, which contained 30,072 loci, was 81.89%. The tissue specific average mapping rate was lowest in rumen, of comparable dimension in jejunum and muscle, and highest in liver.

Long Non-Coding RNA Prediction

Based on the merged annotation, FEELnc predicted 26,740 mRNAs and 7,646 lncRNA transcripts (3,287 loci), out of which 544 were without potential positional interaction partner gene within the default window size of 10,000 to 100,000 nucleotides. Those 7,102 lncRNA transcripts with an assigned potential positional interaction partner were generated by 3,051 loci (**Table 3**, **Supplementary Table 1**). FEELnc distinguishes between intergenic and genic lncRNA with different subtypes (see Wucher et al. (2017) for a graphical explanation). LncRNAs are also classified according to their position to neighboring protein coding genes (interaction partner gene). For intergenic lncRNAs, the best partner gene is closest in terms of distance in base pairs and for genic lncRNAs the best partner gene directly overlaps with it, preferably at an exon. All predicted 7,646 lncRNA transcripts were considered for further computational analyses.

The total of 3,287 lncRNA loci are equally distributed in terms of strandedness (50.6% on the plus strand, 49.41% on the minus strand), and in a locus-based approach (considering the transcript with highest exon number for each locus) the median number of exons per transcript was 3 (average number of exons per transcript: 4.9 ± 8.2). The total exon length geometric mean of the lncRNA loci amounted to 2,201.0 bp.

Prioritized Gene List for Co-Expression Analysis

After filtering the 30,072 genes in the merged annotation for minimal expression (average FPKM across all samples >0.2 in at least one tissue) and exclusion of ribosomal and spliceosomal RNA genes, the dataset contained 22,625 genes out of which 2,886 were lncRNAs, meaning that 401 lncRNAs were removed from RIF and subsequent PCIT co-expression analysis due to very low abundance.

Differential Expression Analysis

The DE analysis yielded a total of 2,154 unique significantly ($q < 0.05$) DE genes between the high and low metabolic efficiency group with 496 DE genes in jejunum, 1,286 DE genes in liver, 479 DE genes in muscle, and no significant differences in rumen (**Figure 2A**). Generally, we observed little overlap of differentially expressed loci between tissues. Out of these unique 2,154 DE genes, 238 were predicted to be lncRNAs corresponding to 11.05%. We observed 40 DE lncRNAs in jejunum, 173 DE lncRNAs in liver, 40 DE lncRNAs in muscle, and none in rumen (**Figure 2B**).

Tissue Enriched Genes

We found a total of 930 genes to be tissue-specifically expressed out of the 22,625 genes, which had passed the initial minimal expression threshold (average expression > 0.2 FPKM in at least on tissue). Out of those 930 genes, 279 were TS in jejunum, 283 in liver, 204 in muscle, and 164 in rumen. Thereof, 21.9% were lncRNAs with 42 in jejunum, 65 in liver, 48 in muscle, and 49 in rumen.

Quantitative Trait Locus Harboring Genes

The database AnimalQTL listed 278 QTL for RFI and 1,881 QTL for milk production traits, which were distributed across 1,615 genes out of which 1,064 passed the minimal expression threshold (average expression > 0.2 FPKM in at least one tissue) in our dataset.

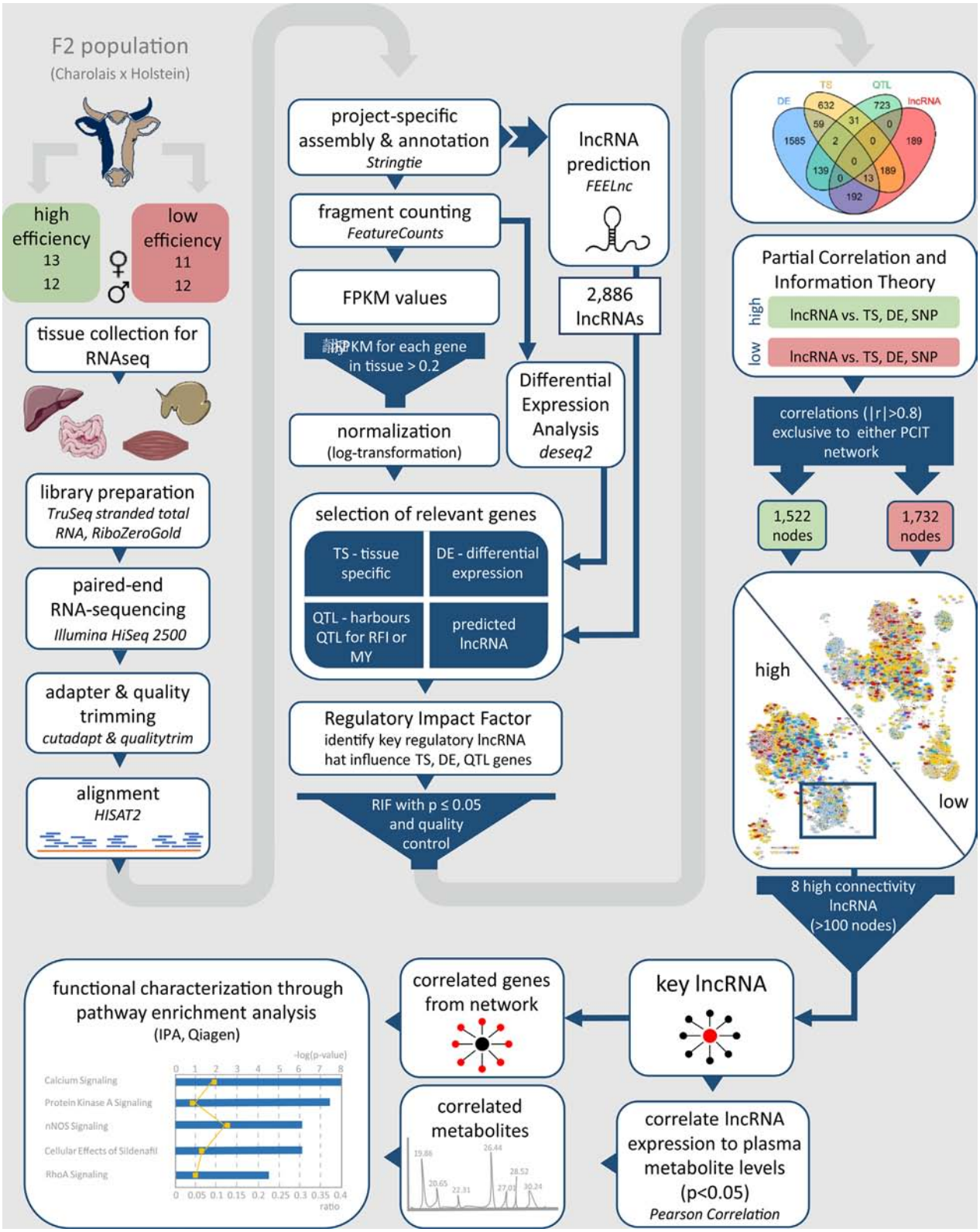


FIGURE 1 | Workflow for the identification and functional characterization of key lncRNAs with regulatory potential in two contrasting biological conditions. The phenotypes under investigation were high and low metabolic efficiency in a Charolais x Holstein cross-population. lncRNA, long non-coding RNA; FPKM, fragments per kilobase transcript length per million reads; TS, tissue specific; DE, differentially expressed; QTL, quantitative trait locus; RFI, residual feed intake; MY, milk production; RIF, regulatory impact factor; PCIT, partial correlation and information theory.

TABLE 2 | Overall and tissue-specific RNA sequencing, alignment, and mapping statistics.

	RIN[1]		Sequencing depth [read pairs]		Alignment to UMD.3.1 [%]		Mapping to project-specific annotation [%]	
	μ[2]	SD[3]	μ	SD	μ	SD	μ	SD
All	8.22	0.81	48,041,209	5,601,638	92.98	9.50	81.89	8.67
Jejunum	8.73	0.44	48,954,376	3,993,201	96.91	0.31	84.99	2.20
Liver	8.00	0.62	50,093,826	5,869,833	98.43	0.20	91.36	1.21
Muscle	7.55	0.85	47,117,156	5,815,843	98.59	0.13	82.42	1.79
Rumen	8.41	0.86	45,999,477	5,587,407	78.00	7.75	69.05	4.67

[1]RIN, RNA integrity number, [2]μ, mean, [3]SD, standard deviation.

Regulatory Impact Factor to Select Long Non-Coding RNAs With a Potential Regulatory Effect on Metabolic Efficiency

The input prioritized gene lists filtered for expression level for the tissue specific RIF analysis contained 2,097 loci for jejunum (880 lncRNAs), 1,890 loci for liver (614 lncRNAs), 961 loci for muscle (363 lncRNAs), and 1,458 loci for rumen (755 lncRNAs). RIF scores were then calculated for the lncRNAs in these gene sets.

With a significance threshold of a RIF1 or RIF2 score ≥ 1.96, the tissue specific RIF analyses identified 92 potential key lncRNAs in jejunum, 55 in liver, 35 in muscle, and 73 in rumen. In total 240 unique lncRNAs had a RIF score ≥ 1.96 in at least one tissue and were considered for subsequent PCIT analysis.

Partial Correlation and Information Theory Approach to Identify Long Non-Coding RNA-Associated Co-Expression Networks

For the within-tissue RIF analysis, the sets of DE genes, TS genes, QTL harboring genes and lncRNAs had been filtered for a seizable expression level (abundance above average expression in the respective tissue) to facilitate a reliable calculation of correlation. For the PCIT analysis, a similar filter for minimal expression was applied: abundance above average expression across all samples in at least one tissue when combining DE genes and TS genes from all tissues with the QTL genes and lncRNAs with significant RIF scores. A total of 295 of the 4,049 prioritized loci were excluded due to not meeting this expression limit. The set of prioritized genes that was used for the final PCIT network analysis contained 3,754 unique genes in total. Thereof, 1,990 were DE genes, 895 QTL containing genes, 926 TS genes, and 583 lncRNAs, though some genes belonged to several categories (**Figure 3**, **Supplementary Table 2**).

The PCIT analysis was performed across tissues and results were filtered for significant correlations with a correlation strength $|r| \geq 0.8$, between a lncRNA with significant RIF score and all genes from the prioritized gene list already used for RIF calculation. Furthermore, correlations had to be exclusive to either the high or low metabolic efficiency group. The high and low network contained 1,522 and 1,732 nodes (genes) respectively (**Supplementary Figure 1**, **Supplementary Figure 2**, **Supplementary Table 3**). Six and two lncRNAs showed a high connectivity (>100 nodes) exclusively in one of the two networks, which represent high and low metabolic efficiency, respectively. Thus, these eight lncRNAs stand out as potential regulatory keys for lncRNAs with respect to metabolic efficiency.

Characterization of Key Regulatory Long Non-Coding RNAs in the Networks

Blast Against New Bovine Assembly

The eight lncRNAs characterized by high connectivity for high and low metabolic efficiency in the PCIT analysis were blasted against the new bovine assembly and annotation [ARS-UCD.1.2, National Center for Biotechnology Information (NCBI) release 106] (**Table 4**). If lncRNAs completely overlapped with annotated genes, the respective lncRNA was located on the opposite strand to the annotated gene (e.g. MSTRG.4926 overlapped with *CDH17* on the opposite strand). None of the eight lncRNA loci had yet been annotated as non-coding in the NCBI or the Ensembl genome annotation (ARS-UCD1.2, release 95).

Pathway Enrichment Analysis

The Pearson correlation analysis between blood plasma metabolites and lncRNA expression, which was calculated prior to the pathway enrichment analysis, showed that the eight key lncRNAs were significantly ($p < 0.05$) correlated to very different numbers of metabolites. Correlations ranged from one (MSTRG.18433) to 117 (MSTRG.4740) metabolites, out of which an average of 75% was successfully mapped in the IPA database and used in the subsequent enrichment analyses (**Supplementary Table 4**). The correlation strength ranged from -0.53 to + 0.48 with an average of |0.35|.

Pathway enrichment analysis for each of the eight key lncRNAs with their respective correlated metabolites and genes showed that calcium signaling was the most strongly enriched canonical pathway for half of the key lncRNAs (MSTRG.9051, MSTRG.10337, MSTRG.18433, and MSTRG.19312). The other high ranking canonical pathway hits, i.e. hits with the lowest p-value, were tRNA charging, leukocyte extravasation signaling, caveolar-mediated endocytosis signaling, and T cell receptor signaling (data not shown).

Within the eight lncRNAs with a high connectivity in the PCIT analysis, three loci showed distinct pattern in the pathway enrichment analysis suggesting divergent molecular functions. Inspection of the results showed that the enriched canonical pathways for MSTRG.4740, which was differentially expressed in

TABLE 3 | Characterization of high connectivity long non-coding RNAs from networks specific for high or low metabolic efficiency in cattle.

Identifier		FEElnc Prediction (based on UMD3.1 release 92)			FPKM		Position & Structure			Differential Expression Analysis		
MSTRG	**TN**	**Closest partner gene (gene symbol)**	**Direction, type, location**	**Distance bp**	**Mean**	**BTA**	**Exons**	**Start bp**	**Strand**	**Exonic length**	**Log2 FC**	**p-value**	**q-value (BH)**	**Tissue**
MSTRG.4740	1	ENSBTAG00000002062 (TRPA1)	Anti-sense, genic, intronic	0	6.74	14	2	37760114	+	437	0.77	2.10E-04	9.13E-03	Liver
MSTRG.4926	1	ENSBTAG00000021964 (CDH17)	Anti-sense, genic, exonic	0	0.78	14	18	72437085	-	3,321	NA	NA	NA	None
MSTRG.9051	1	ENSBTAG00000004651 (NME1)	Anti-sense, intergenic, downstream	358	2.57	19	2	36225984	-	1,866	NA	NA	NA	None
	2			1,587				36227213						
MSTRG.10337	1	ENSBTAG00000005353 (DES)	Anti-sense, genic, exonic	0	5.37	2	9	108075380	-	2,989	NA	NA	NA	None
MSTRG.17681	1	ENSBTAG00000005726 (HNRNPA2B1)	Anti-sense, intergenic, upstream	9,170	19.80	4	3	70184682	-	1,552	-0.38	8.45E-05	5.04E-03	Liver
MSTRG.18433	1	ENSBTAG00000015828 (FKBP11)	Sense, intergenic, upstream	5,808	8.94	5	2	31038376	+	2,683	NA	NA	NA	None
MSTRG.19098	1	ENSBTAG00000046324 (C-type lectin domain family 2 member D11)	Anti-sense, genic, exonic	0	1.72	5	5	100926473	+	1,570	-0.63	1.27E-06	3.04E-04	Liver
MSTRG.19312	1	ENSBTAG00000009886 (KDELR3)	Anti-sense, genic, exonic	0	25.84	5	3	110665971	-	5,768	NA	NA	NA	None
	3							110670286						

**MSTRG, identifier from project-specific bovine transcriptome annotation; TN, transcript number; FPKM, fragments per kilobase million; BTA, bovine chromosome; bp, base pair; FC, foldchange; BH, Benjamini–Hochberg, NA, not available.*

FIGURE 2 | Venn diagrams of **(A)** all loci **(B)** exclusively lncRNAs with differential expression (DE) between high and low metabolic efficiency in cattle. DE analysis was performed within the tissues jejunum, liver, muscle, and rumen. No loci were significantly [q-value (Benjamini–Hochberg) < 0.05] DE in rumen.

liver (**Figure 4**, **Table 3**, **Supplementary Table 5**), were related to amino acid biosynthesis and metabolism, as well as protein synthesis (**Table 5**). MSTRG.17681 (**Figure 5**, **Supplementary Table 5**) which was also differentially expressed in liver, seemed to act very locally in the coatomer subunit of the coat protein I (COPI) in the caveosome. MSTRG.10337, (**Figure 6**, **Supplementary Table 5**) apparently acts specifically in muscle where it was related to several signaling pathways, most strongly to calcium, protein kinase A, neuronal nitric oxide synthase (nNOS), and RhoA signaling (**Table 5**).

DISCUSSION

A major goal of this study was the identification of lncRNAs that hold a potential key regulatory role in metabolic efficiency,

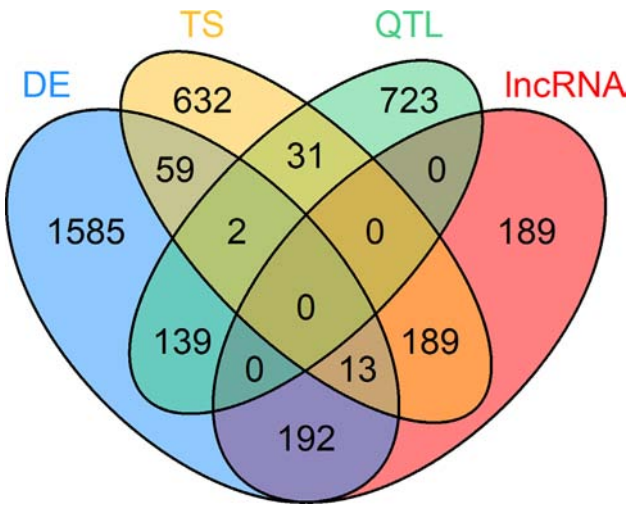


FIGURE 3 | Venn diagram of 3,754 loci selected for co-expression network construction. Loci belonging to at least one of these four categories were considered: differential expression (DE) in at least one tissue, tissue specific (TS) expression, harboring a QTL for residual feed intake and or milk production (QTL) and key regulatory long non-coding (lnc) RNAs [significant ($p < 0.05$) regulatory impact factor score].

which was roughly defined as the animal's ability to direct the energy adsorbed into protein synthesis and use it for muscle mass accumulation or milk secretion. We integrated phenotypic, metabolomics and transcriptomics data from a cattle F_2-population (Charolais × Holstein) in a co-expression network approach to mine for lncRNAs with a regulatory role in metabolic processes. By contrasting animals of high and low metabolic efficiency and by including RNAseq data from four key metabolic tissues in a combined analysis, we identified highly connected hub lncRNAs. Finally, we subjected metabolites and genes, whose plasma levels or transcript abundance significantly correlated with expression levels of the specific, highly connected lncRNA, to the integrative approach for metabolomics and transcriptomics data as offered by the cross-platform IPA (Kramer et al., 2014).

Establishment of a Pipeline Based on Regulatory Impact Factor and Partial Correlation and Information Theory to Establish Co-Expression Networks for Long Non-Coding RNAs and Genes to Predict Their Role in Metabolic Efficiency

Weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) is a frequently applied method to identify co-expression pattern at whole transcriptome level. Recently, Sun et al. (2019) applied this method for mining regulatory signatures of divergent feed efficiency in beef cattle investigating a multi-tissue transcriptome data set. WGCNA has also been used to find hub lncRNAs in a transcriptomic landscape in multiple studies in humans as well as animals (Miao et al., 2016; Tang et al., 2017; Li et al., 2018; Weikard et al., 2018; Wang et al., 2019). To mine for the functional role of lncRNAs of interest via WGCNA, one might select lncRNAs that are strongly correlated with coding neighbor genes (Li et al., 2018) or lncRNAs that were differentially expressed between conditions or phenotypes (Weikard et al., 2018; Wang et al., 2019).

TABLE 4 | BLAST results for eight high connectivity long non-coding RNAs (>100 nodes) in partial correlation and information theory networks with connections exclusive for high or low metabolic efficiency.

lncRNA		BLAST against bovine reference genome (ARS-UCD1.2, release 106)				
Identifier	**Network (connectivity in nodes)**	**Annotated gene with highest sequence homology**	**Identity [%]**	**Query cover [%]**	**E-Value**	**Position of lncRNA relative to homologous gene in ARS-UCD1.2**
MSTRG.4740	Low (147)	mRNA-transient receptor potential cation channel subfamily A member 1 (TRPA1)	100.00	100.00	9.00E-116	Intronic, anti-sense
		ADP-ribosylation factor 4 (ARF4)	98.57	91.00	3.00E-100	Exonic, sense
MSTRG.4926	High (144)	Cadherin-17 precursor (CDH17)	100.00	100.00	0.00E+00	Anti-sense
MSTRG.9051	High (170)	Nucleoside diphosphate kinase A 1 isoform X1 (NME1)	99.72	100.00	0.00E+00	Sense, genic
MSTRG.10337	Low (239)	Desmin (DES)	99.93	100.00	0.00E+00	Exonic, anti-sense
MSTRG.17681	High (120)	39,201 bp at 5' side: alpha-aminoadipic semialdehyde synthase, mitochondrial precursor 88559 bp at 3' side: fez family zinc finger protein 1	98.40	99.00	0.00E+00	Sense, genic
		Chromobox protein homolog 3 isoform X1 (CBX3)	99.00	89.00	0.00E+00	Sense, genic
MSTRG.18433	High (268)	364 bp at 5' side: ADP-ribosylation factor 3; 37831 bp at 3' side: peptidyl-prolyl cis-trans isomerase FKBP11 precursor	99.96	100.00	0.00E+00	Sense, intergenic
MSTRG.19098	High (184)	C-type lectin domain family 2 member D11	100.00	100.00	0.00E+00	Anti-sense, genic
MSTRG.19312	High (212)	ER lumen protein-retaining receptor 3 (KDELR3)	100.00	99.00	0.00E+00	Anti-sense, genic

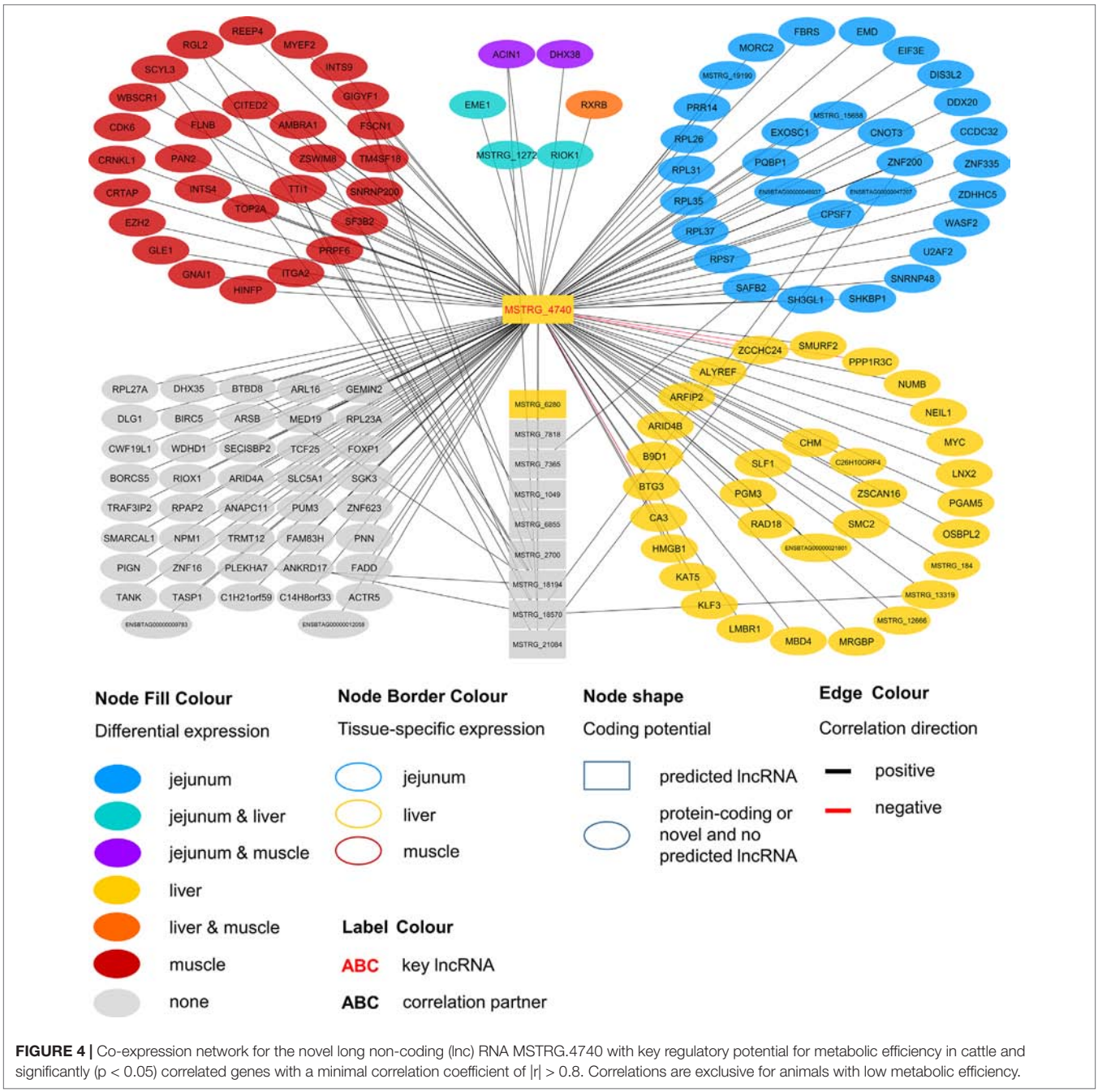


FIGURE 4 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.4740 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with low metabolic efficiency.

The connectivity within a network and the differential wiring between two networks can also serve as a selection criterion (Pellegrina et al., 2017). In our study we present an alternative approach for the selection of lncRNAs of interest, the RIF (Reverter et al., 2010), which has already successfully been applied to transcription factors (TF). In combination with a PCIT (Reverter and Chan, 2008), key regulatory TFs during puberty could be identified in cattle (Cánovas et al., 2014), as well as critical TFs in porcine muscle (Perez-Montarelo et al., 2012). This approach seemed to be particularly applicable for lncRNAs with regard to the expression level as they generally exhibit lower transcript abundance compared with mRNAs (Derrien et al., 2012), as do TFs compared with other coding genes (reviewed by Vaquerizas et al., 2009). We indeed found that only 10% of the unique lncRNAs with a significant RIF-score (n = 240) were also differentially expressed, including three of the eight key hub lncRNAs. LncRNAs were significantly underrepresented in the list of DE loci across all tissues (X^2 test, p = 1.2E-06): while they accounted for 14.85% of all loci in the DE analyses, only 11.05% of the DE loci were classified as lncRNAs. In contrast, the other

TABLE 5 | Top 10 enriched pathways derived from genes and metabolites significantly correlated with key long non-coding RNAs associated with metabolic efficiency

ID	Ingenuity Canonical Pathways	log(p)	Ratio	Molecules
MSTRG.4740	tRNA Charging	5.56E00	8.54E-02	L-valine, L-phenylalanine, L-tryptophan, glycine, L-arginine, L-tyrosine, L-lysine
	EIF2 Signaling	4.13E00	3.83E-02	MYC, RPS7, RPL27A, RPL35, RPL23A, RPL37, RPL26, EIF3E, RPL31
	Glucose and Glucose-1-phosphate Degradation	3.18E00	1.3E-01	D-glucose, PGM3, phosphate
	Tyrosine Biosynthesis IV	2.94E00	2.86E-01	L-phenylalanine, L-tyrosine
	Acetyl-CoA Biosynthesis III (from Citrate)	2.82E00	2.5E-01	phosphate, citric acid
	Glycine Degradation (Creatine Biosynthesis)	2.71E00	2.22E-01	glycine, L-arginine
	Phenylalanine Degradation IV (Mammalian, via Side Chain)	2.68E00	8.82E-02	L-phenylalanine, phenylpyruvic acid, glycine
	Glutathione Biosynthesis	2.53E00	1.82E-01	phosphate, glycine
	Thymine Degradation	2.53E00	1.82E-01	5, 6-dihydrothymine, beta-ureidoisobutyric acid
MSTRG.10337	Calcium Signaling	1.63E01	9.35E-02	TNNT1, CHRNA1, CACNB1, CACNG1, CACNA1S, MYL2, TNNI2, TNNT3,T NNC2, TNNC1, MYL1, ATP2A1, CAMK2A, CASQ1, RYR1, TNNI1, CASQ2, MYL3, ACTA1, CAMK2B
	Protein Kinase A Signaling	7.45E00	3.88E-02	TNNI2, MYL2, MYLPF, MYLK2, PPP1R3A, TTN, MYL1, EPM2A, CAMK2A, PLCB1, RYR1, TNNI1, EYA1, MYL3, CAMK2B, PHKG1
	nNOS Signaling in Skeletal Muscle Cells	6.1E00	1.3E-01	CACNG1, CACNB1, CACNA1S, CHRNA1, RYR1, L-arginine
	Cellular Effects of Sildenafil (Viagra)	6.09E00	6.25E-02	CACNA1S, CACNG1, MYL2, MYLPF, PLCB1, L-arginine, MYL1, MYL3, ACTA1
	RhoA Signaling	4.55E00	5.6E-02	MYL2, MYLPF, MYLK2, TTN, MYL1, MYL3, ACTA1
	Apelin Cardiomyocyte Signaling Pathway	3.7E00	5.00E-02	MYL2, MYLPF, PLCB1, MYL3, MYL1, ATP2A1
	Actin Cytoskeleton Signaling	3.55E00	3.36E-02	MYL2, MYLPF, ACTN3, MYLK2, TTN, ACTA1, MYL3, MYL1
	Regulation of Actin-based Motility by Rho	3.24E00	5.21E-02	MYL2, MYLPF, MYL3, ACTA1, MYL1
	ILK Signaling	3.19E00	3.38E-02	PARVB, MYL2, TNFRSF1A, ACTN3, MYL1, MYL3, ACTA1
	Thrombin Signaling	2.93E00	3.06E-02	CAMK2A, MYL2, MYLPF, PLCB1, MYL1, MYL3, CAMK2B
MSTRG.17681	Caveolar-mediated Endocytosis Signaling	3.56E00	5.48E-02	ARCN1, COPA, COPE, COPB2
	Fatty Acid α-oxidation	2.29E00	8.00E-02	ALDH3A2, ALDH9A1
	Death Receptor Signaling	2.15E00	3.3E-02	PARP10, PARP4, HTRA2
	Histamine Degradation	2.05E00	6.06E-02	ALDH3A2, ALDH9A1
	Oxidative Ethanol Degradation III	2.05E00	6.06E-02	ALDH3A2, ALDH9A1
	G Protein Signaling Mediated by Tubby	2.03E00	5.88E-02	GNG2, GNAQ
	Tryptophan Degradation X (Mammalian, via Tryptamine)	2.00E00	5.71E-02	ALDH3A2, ALDH9A1
	Putrescine Degradation III	2.00E00	5.71E-02	ALDH3A2, ALDH9A1
	Ethanol Degradation IV	1.98E00	5.56E-02	ALDH3A2, ALDH9A1
	NER Pathway	1.96E00	2.8E-02	HIST2H4B, XAB2, RAD23B

loci accounted for 85.25% of all loci in the DE analyses, but had a share of 88.95% in the total of 2,154 differentially expressed unique loci.

In a recent publication, van Dam et al. (2017) reviewed and highlighted the usefulness of gene co-expression networks for the functional classification of genes and novel loci, such as non-coding elements without any known function. Correspondingly Oliveira et al. (2018) successfully applied a co-expression network concept to identify genes and miRNAs regulating IMF in Nellore steers. Besides the preselection of lncRNAs for co-expression networks, it might be advisable to make a knowledge-based preselection also for other genes to be included instead of simply using all expressed genes. The combination of RNA-Seq results with GWAS hits (gene regions associated with QTL for milk performance traits or RFI) is an acknowledged procedure to integrate multiple layers of knowledge into a prioritized gene set for co-expression network analysis (Schaefer et al., 2018). In our PCIT analysis, we prioritized genes that appeared to be functionally important from the RNA-Seq analysis [DE loci (2,154) or TS loci (930)] and published GWAS data and selected those for our prioritized gene set to create a stronger focus on bovine metabolic efficiency, accepting however that still unknown, yet important elements might be overlooked. When preparing the prioritized gene set, we noted that the key role of liver in metabolic processes was clearly reflected by the by far highest number of DE loci (1,286) between the two metabolic

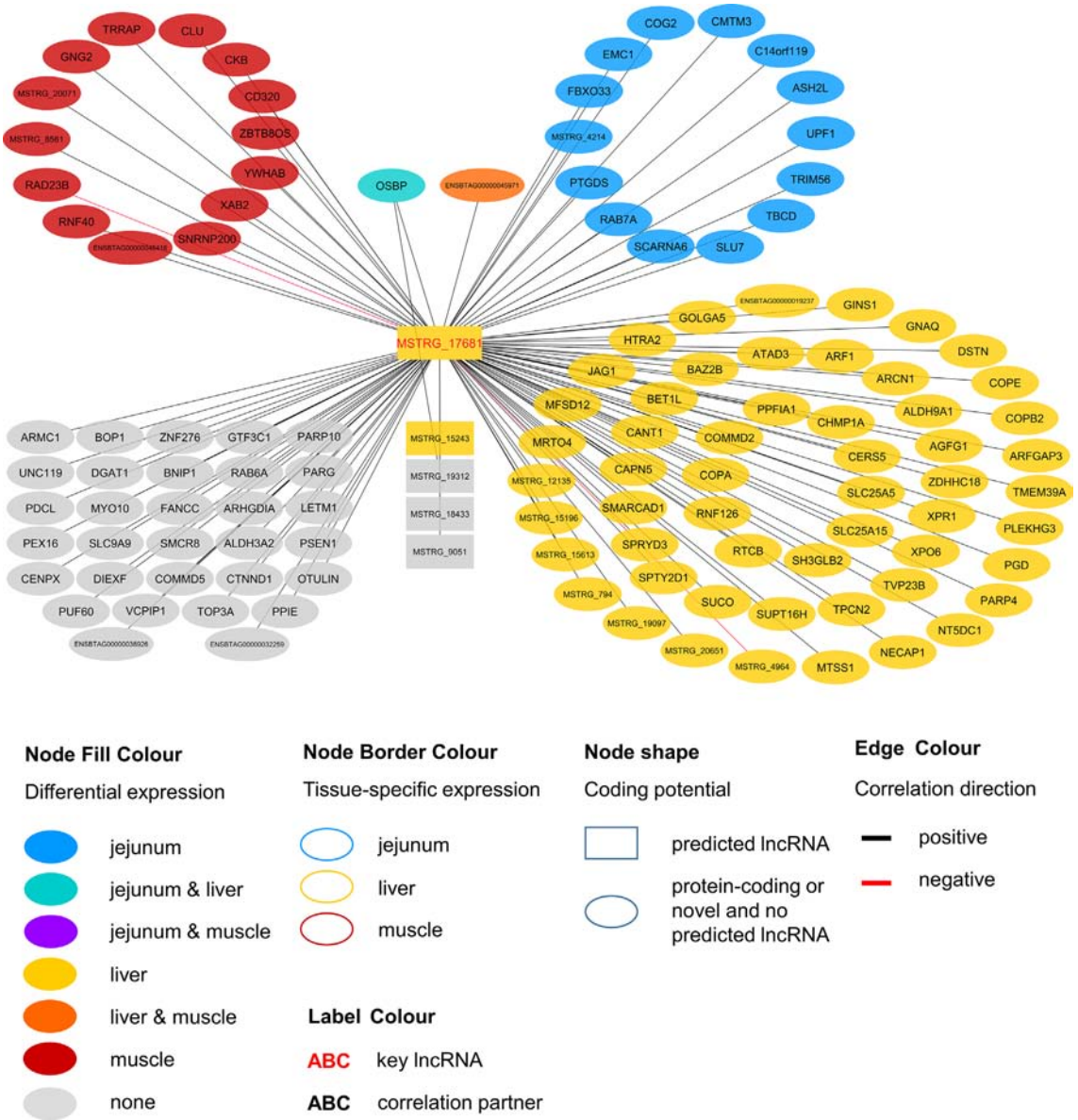


FIGURE 5 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.17681 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with high metabolic efficiency.

efficiency groups, which was 2.6 fold higher than in jejunum or rumen. For DE loci in the prioritized gene set that was used for the PCIT, we noted that these predominantly (65%) had their highest expression in a different tissue than where they were differently expressed. This underlines that tissue specificity or tissue of highest abundance and DE of loci are indeed different, non-redundant features and that it is recommendable to follow a TS perspective in the beginning of the analysis.

One way to deduce a biological function of lncRNAs is to take a close look at coding genes in their immediate vicinity. This idea has also been implemented in the bioinformatics tool FEELnc for lncRNA prediction and annotation (Wucher et al., 2017), where the potential partner gene is generally assumed to be the closest annotated gene. However, this exclusively focusses on *in-cis* interaction with a narrow frame of impact. However, it has been reported that some lncRNAs execute *in-trans* regulatory tasks by binding directly to distant DNA sites or via RNA-protein interactions (Long et al., 2017) or a direct effect on RNA polymerase II activity (Kornienko et al., 2013).

Another way to infer functionality of unknown genomic elements subsequent to the network construction is to submit correlated coding genes to an enrichment analysis (Chen et al., 2018b), thereby assuming the guilt-by-association principle. Following this approach, we took genes from the prioritized

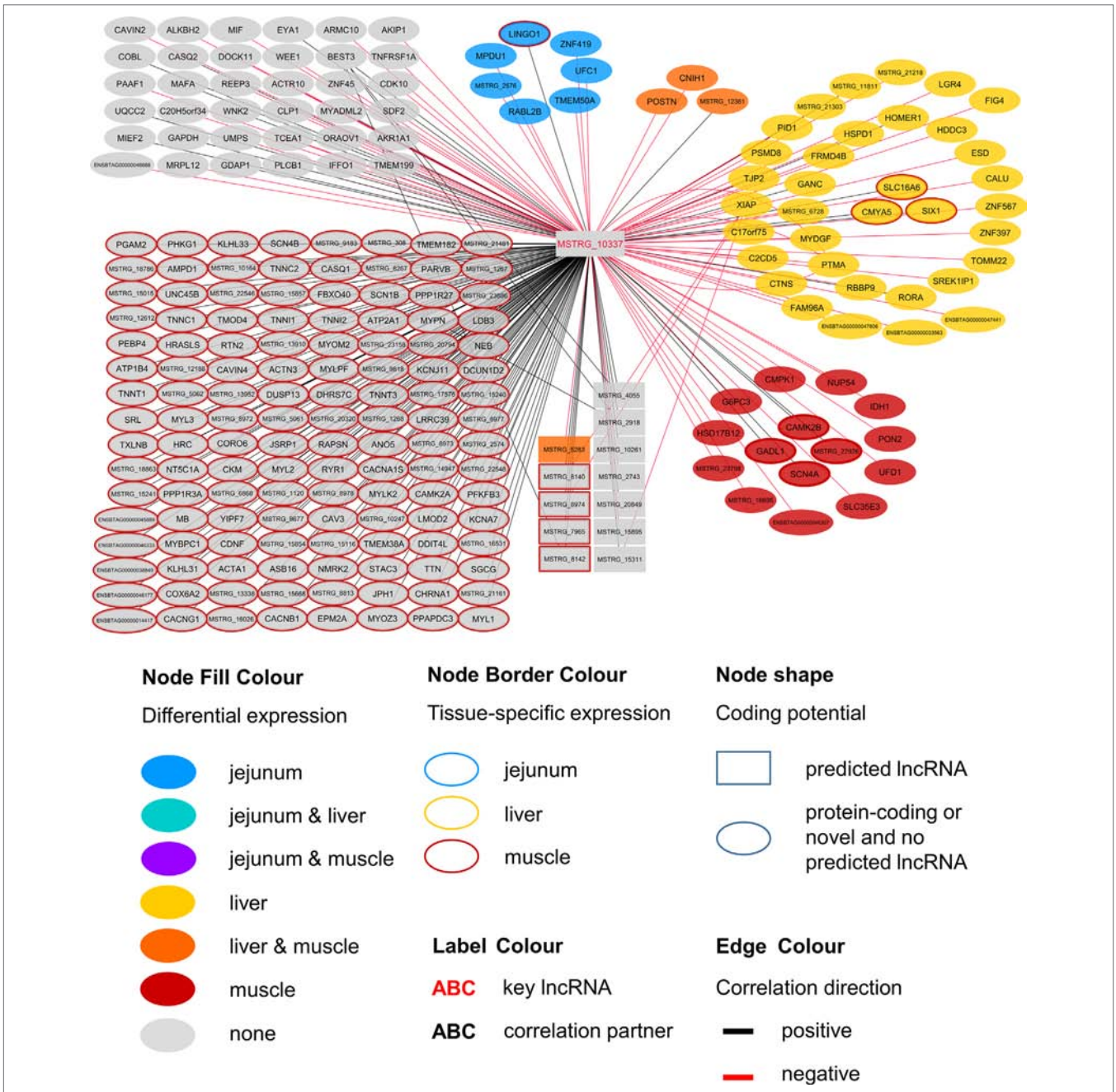


FIGURE 6 | Co-expression network for the novel long non-coding (lnc) RNA MSTRG.10337 with key regulatory potential for metabolic efficiency in cattle and significantly ($p < 0.05$) correlated genes with a minimal correlation coefficient of $|r| > 0.8$. Correlations are exclusive for animals with low metabolic efficiency.

gene set that were correlated with high connectivity lncRNAs of interest. LncRNA partner genes predicted by FEELnc could also be part of the prioritized gene set if they fell into one of the categories (DE, tissue-specificity, QTL-harboring). This was the case for 473 out of 2,741 unique predicted lncRNA interaction partner genes. Thus, 12.6% of the genes that were used as PCIT input (3,754) were very close to or overlapped with a lncRNA.

In addition, we aimed to add a supplementary layer of information to the pathway enrichment analysis and thereby to create further biological depth by using the option to integrate gene expression and metabolic profiles. In a single step this approach facilitates to predict a link between transcriptome activity, the direct functional readout of metabolic activity or physiological status and the functional analysis of lncRNAs. MSTRG.4740, e.g., correlated with plasma levels of 117 metabolites—valuable information that would otherwise be missing from the enrichment analysis. To our knowledge, we here present the first study that integrates metabolomics and

transcriptomic data in an enrichment analysis to predict the functional role of lncRNAs.

Across-Tissue Candidate Long Non-Coding RNAs for Metabolic Efficiency

LncRNAs were defined as hubs when they were connected to at least 100 other nodes in the high or low efficiency PCIT network. Three of the identified eight hub lncRNAs were exemplarily chosen for a more detailed description of their biological functionality predicted with IPA. These lncRNAs—namely MSTRG.4740, MSTRG.10337, and MSTRG.17681—were hubs in gene groups that showed enrichment for transfer RNA (tRNA) charging (p = 2.78E-06) and EIF2 signaling (p = 7.34E-05), calcium signaling (p = 4.98E-17) and nNOS signaling in skeletal muscle cells (p = 7.88E-07), and calveolar-mediated endocytosis signaling (p = 2.77E-04) and fatty acid oxidation (p = 5.13E-03), respectively.

For MSTRG.4740 an encompassing look at the enriched pathways clearly pointed towards amino acid metabolism and protein synthesis. This lncRNA was DE in liver (adjusted p-value (BH) = 9.13E-03, log2FC = 1.70) but displayed highest abundance (average FPKM) in jejunum (10.68) and rumen (8.41) and lowest in muscle (1.66) compared to liver (6.23). The DE status in liver suggested biological relevance there. However, the RIF analysis attributed a significant score to MSTRG.4740 in jejunum. The strongest enrichment was for tRNA charging (p = 2.78E-06), which describes the attachment of amino acids to a tRNA before incorporation into a growing polypeptide. According to IPA, the enrichment of this pathway was due to the correlation of MSTRG.4740 expression level with the blood plasma content of six essential or semi-essential amino acids (L-valine, L-phenylalanine, L-tryptophan, L-arginine, L-tyrosine, L-lysine). No non-essential amino acid showed a significant correlation with this lncRNA. The significantly correlated amino acids play integral roles as regulators of metabolism and key body functions, but cannot or only partially be synthesized by bovine animals themselves. Plasma concentration of essential amino acids depends on uptake from the diet, the balance between protein synthesis and degradation in peripheral tissues as well as on the efficiency of transport processes. The enrichment of the tRNA Charging pathway was not backed up by other components in addition to the indicated amino acids (e.g., charged tRNAs themselves). Thus, we restrict our conclusion and suggest that the lncRNA has a close relationship with (semi-) essential amino acid levels, but rather not to tRNA Charging per se. Widmann et al. (2015) reported no significant correlation between plasma amino acids and RFI at the onset of puberty in bulls in the same resource population. However, in the current study we employed adult animals.

Endogenous metabolism and also supply of amino acid have been demonstrated to limit growth or lactation in pigs, cattle and fish as reviewed by Hou et al. (2016). Furthermore, Doelman et al. (2015) showed that an abomasal infusion with essential amino acids leads to increased protein levels of eIF2α and eIF2Bε in the mammary gland in dairy cows. The authors proclaimed a direct link between the eIF2 factor, which is essential for eukaryotic translation initiation and milk protein yield. Interestingly, we found *eIF2Bε* to be DE [q-value (BH) = 0.022, log2FC = 0.204] in liver and to be one of the genes underlying the significant enrichment of the EIF2 Signaling pathway (p = 7.34E-05), which is tightly linked to protein synthesis. Genes encoding for ribosomal proteins of 40S (*RPS7*) or 60S subunits (e.g. *RPL26*, *RPL31*) were significantly correlated with MSTRG.4740, as well as the before mentioned *eIF2Bε*. EIF2 signaling and subsequently *EIF3E* are required for the correct initiation of mRNA translation (Kimball 1999; Walsh and Mohr, 2014).

Considering the presented correlations of MSTRG.4740 with other genes and plasma metabolites, this hub lncRNA seems to be an excellent example of a potential new key regulator in metabolic efficiency through the modulation of translational processes.

In contrast to MSTRG.4740 that seems to act on the broader forefront of translation, MSTRG.17681 appears to have a rather narrow and more targeted function. The first hit in pathway enrichment was calveolar-mediated endocytosis signaling (p = 2.77E-04). Four genes (*COPA*, *COPE*, *COPB2*, *ARCN1*) belonging to this pathway were highly correlated ($|r| > 0.8$) with this hub lncRNA. We observed significant DE in the liver of divergently efficient animals for MSTRG.17681 (q-value (BH) = 0.0050, log2FC = 0.766) as well as the respective quartet of genes. *COPA*, *COPE* and *COPB2* are transporters and *ARCN1* encodes the coatomer subunit of the coat protein I (COPI) complex (Tunnacliffe et al., 1996). All genes are allocated to a subunit in the cellular calveolar-mediated endocytosis signaling: the COPI vesicle, which plays a role in intracellular lipid transport (Popoff et al., 2011) and regulates lipid homeostasis (Beller et al., 2008). COPI-vesicle biogenesis is *ARF1*-dependent (Beck et al., 2009), which we found to be DE in liver and to be positively correlated with MSTRG.17681. The *Arf1 GTPase-activating protein* 3 (*ArfGAP3*) that subsequently allows the vesicle to fuse with a target membrane (Beck et al., 2009), was also correlated to MSTRG.17681 and DE in liver.

Considering that COPI-vesicles assist in lipid transport, it seems fitting that we found significant correlations between MSTRG.17681 expression and plasma levels of two saturated fatty acids: caprylate (p = 0.013, r = 0.357) and heptanoate (p = 0.047, r = 0.289). Caprylic acid supplementation in the diet of weaned piglets was observed to lead to a significant increase body weight gain (Marounek et al., 2004). MSTRG.17681 most likely acts predominantly in jejunum, liver, and rumen, where average expression was much higher (31.83, 25.26, and 18.74 FPKM, respectively) compared with the expression in skeletal muscle (3.36 FPKM). We infer that MSTRG.17681 is a key regulator in COPI-vesicle functioning and thereby presumably affects lipid levels.

MSTRG.10337 was the third key hub lncRNA with a distinct prediction of biological function. In the network specific for animals of low metabolic efficiency, MSTRG.10337 was co-expressed with 39 genes that were DE in liver, 4 of which were also DE in muscle. Interestingly, the hub lncRNA MSTRG.10337 correlated with *RORA* (*RAR related orphan receptor A*), which was DE in liver. *RORA* is a transcriptional regulator of genes related to lipid metabolism, e.g. *APOA1*, *APOA5*, *APOC3*, and *PRAPRG* (Vu-Dac et al., 1997; Raspe et al., 2001; Sundvold and Lien, 2001; Lind et al., 2005). Although not meeting the threshold

for entering the PCIT network with respect to correlation to MSTRG.10337, we found *APOA1* to be DE in the liver, providing consistency in gene expression and biological interplay with regard to *RORA*. Previously, Krappmann et al. (2012) has attested an association of a *RORC* (*RAR Related Orphan Receptor C*) variant with milk yield, as well as milk fat and protein percentage in our SEGFAM resource population. Furthermore, Zhang et al. (2017) linked both nuclear receptors *RORA* and *RORC* to hepatic lipid and fatty acid metabolism as well as circadian rhythm pathways in a liver-specific depletion experiment in mice.

The most enriched pathways related to MSTRG.10337 are Calcium signaling (p = 4.98E-17) Protein Kinase A (PKA) signaling (p = 3.51E-08), and nNOS signaling in skeletal muscle cells (p = 7.88E-07). These data confirmed findings from an alternative previous network analysis in our resource population, where GWAS results for RFI and metabolomics profiles were merged for bulls in puberty. Widmann et al. (2015) also has identified Protein Kinase A (PKA) signaling and Nitric Oxide signaling to be significantly enriched pathways in IPA analyses.

Calcium signaling, Protein Kinase A (PKA) signaling and nNOS signaling in skeletal muscle cells are in biological interplay. Protein kinases are in charge of nNOS phosphorylation on different serine residues and catalyze the hydroxylation of L-arginine (Fleming, 2008). In turn, L-arginine plasma levels were negatively correlated with expression levels of MSTRG.10337 (p=0.038, r=-0.323) in our study. This would fit an inhibitory role of MSTRG.10337 in metabolic efficiency, because of unfavorable effects of arginine depletion in the diet on milk protein synthesis in dairy cows (Tian et al., 2017). The inhibitory effect is underlined by numerous negative correlations of MSTRG.10337 to genes with DE in liver (e.g. *LGR4*, *FIG4*, *ESD*), muscle (e.g. *PON2*, *IDH1*, *NUP54*) and jejunum (e.g. *LINGO1*, *MPDU1*, *UFC1*), as well as QTL harboring genes (e.g. *GAPDH*, *MAFA*, *MYBPC1*), although the exact mode of operation is unclear. The supplementation of arginine has been reported to reduce body fat deposition, improve muscle gain and improve insulin sensitivity and the metabolic profile (Wu et al., 2009), and its availability in the organism is therefore particularly interesting for beef production. In chicken, L-arginine supplementation enhanced lean muscle growth (Castro et al., 2018). However, protein anabolic effects in muscle via dietary arginine supplementation are controversially discussed in other species (Tang et al., 2011). In addition to Calcium and PKA signaling, a third highly enriched pathway for MSTRG.10337 was nNOS signaling. In terms of gene expression, nNOS is not restricted to neuronal cells but is commonly expressed in skeletal muscle and certain vascular smooth muscle cells as well (Fleming 2008), where it is important for tissue integrity and contractile performance (Percival, 2011). After Ca^{2+}-activation, nNOS enzymes produce NO, which affects the autoregulation of blood flow, myocyte differentiation and glucose homeostasis in skeletal muscle cells (Stamler and Meissner, 2001). In a previous study we already suspected a relationship between NO signaling, arginine and growth in cattle (Widmann et al., 2013).

We assume that MSTRG.10337 influences the onset of nNOS activation, because of its correlation to calcium voltage-gated channel genes and *RYR1* (*ryanodine receptor 1*) that encodes a calcium release channel protein (Loy et al., 2011). Co-expression with a large number of muscle specific genes (e.g. *CACNG1, MYLK2, TNNT1, MYL2*) or genes that are DE in muscle (*CAMK2B*) related this hub lncRNA to PKA and nNOS signaling. It might thereby influence phosphorylation, degradation and availability of L-arginine in the muscle cells, but simultaneously perform some regulatory tasks in hepatic lipid metabolism.

CONCLUSIONS

In this study, we were able to identify novel lncRNAs with potential key regulatory function in metabolic efficiency in cattle. Although usually low expression levels of lncRNAs entail difficulties in DE and co-expression analyses, the careful setting of expression thresholds, the use of a-priori knowledge in gene prioritization and the integrated use of RIF metrics and PCIT based co-expression networks have proven to be a valid method for the identification of regulatory hub lncRNAs. The enrichment analysis based on metabolites and gene expression data provided valuable insight into the putative biological functions of yet uncharacterized lncRNAs.

We focused on phenotypic differences and looked at mechanisms or correlations that were exclusive to either metabolic efficiency group. Still, other correlations between lncRNAs and mRNAs might exist simultaneously in both groups, and we propose to take a group transcending approach in a follow-up study. For future work, we suggest to proceed within tissues to get a clearer picture of gene-gene interactions within a tissue, also because we noted that a multi-tissue approach presents its challenges when interpreting pathway enrichment results. The hub lncRNAs, which we identified, can be considered as candidates for further validation studies, *in vitro* or *in vivo*. Kashi et al. (2016) neatly described modern methods to determine where and how lncRNAs act in the cell or organism, such as chromatin isolation by RNA purification (ChIRP) sequencing (Chu et al., 2011).

In conclusion, our study demonstrates that the method we presented is suitable for the identification for key regulatory lncRNAs in a complex phenotype. By carefully adjusting different elements of the procedure, e.g. the tissue under consideration or the choice of priority categories for genes to include in the network analysis, this pipeline allows us to answer targeted biological questions.

DATA AVAILABILITY STATEMENT

The datasets generated for this study have been submitted to the "Functional Annotation of Animal Genomes" (FAANG) initiative database, accession PRJEB34570, and are also available via the European Nucleotide Archive (ENA).

ETHICS STATEMENT

The animal study was reviewed and approved by Animal care and experimental procedures following the guidelines of the German Law of Animal Protection. The protocols were approved by the Animal Protection Board of the Leibniz Institute for Farm Animal Biology as well as by the Animal Care Committee of the State Mecklenburg-Western Pomerania, Germany (State Office

for Agriculture, Food Safety and Fishery; LALLF M-V/ Rostock, Germany, TSD/7221.3-2.1-010/03).

AUTHOR CONTRIBUTIONS

WN performed the statistical analyses and investigations, created the visualizations and wrote the original draft. RW and CK performed data collection, generated transcriptomic data, contributed to data analysis and conceptualized and administered the project and supervised WN. AR coded and performed bioinformatics analyses and supervised WN. RB, EA, and HH provided support with sampling and phenotyping of the test animals. All authors contributed to reviewing and editing the manuscript.

FUNDING

This study was funded by the German Research Foundation (DFG–grant numbers: KU 771/8-1 and WE 1786/5-1). WN received a scholarship for doctoral candidates from the German Academic Exchange Service (DAAD) and travel funds from the Graduate Academy of the University of Rostock. The publication of this article was funded by the Open Access Fund of the Leibniz Institute for Farm Animal Biology (FBN).

ACKNOWLEDGMENTS

The authors thank Frieder Hadlich for his support with bioinformatics obstacles of all kind, Marina Naval-Sanchez for her insightful ideas in network analysis, and Simone Wöhl and Bärbel Pletz for their excellent technical work in the lab.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01130/full#supplementary-material

REFERENCES

Alexandre, P. A., Naval-Sanchez, M., Porto-Neto, L. R., Ferraz, J. B. S., Reverter, A., and Fukumasu, H. (2019). Systems biology reveals NR2F6 and TGFB1 as key regulators of feed efficiency in beef cattle. *Front. Genet.* 10, 230. doi: 10.3389/fgene.2019.00230

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrew, S. (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Archer, J. A., Arthur, P. F., Herd, R. M., Parnell, P. F., and Pitchford, W. S. (1997). Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *J. Anim. Sci.* 75, 2024–2032. doi: 10.2527/1997.7582024x

Arnes, L., Akerman, I., Balderes, D. A., Ferrer, J., and Sussel, L. (2016). Betalinc1 encodes a long noncoding RNA that regulates islet beta-cell formation and function. *Genes Dev.* 30, 502–507. doi: 10.1101/gad.273821.115

Beck, R., Ravet, M., Wieland, F. T., and Cassel, D. (2009). The COPI system: molecular mechanisms and function. *FEBS Lett.* 583, 2701–2709. doi: 10.1016/j.febslet.2009.07.032

Beller, M., Sztalryd, C., Southall, N., Bell, M., Jäckle, H., Auld, D. S., et al. (2008). COPI Complex is a regulator of lipid homeostasis. *PloS Biol.* 6, e292. doi: 10.1371/journal.pbio.0060292

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., et al. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* 50, 362–367. doi: 10.1038/s41588-018-0056-5

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J. B., et al. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542. doi: 10.1016/0092-8674(92)90520-M

Burkard, C., Opriessnig, T., Mileham, A. J., Stadejek, T., Ait-Ali, T., Lillico, S. G., et al. (2018). Pigs lacking the scavenger receptor cysteine-rich domain 5 of CD163 are resistant to porcine reproductive and respiratory syndrome virus 1 infection. *J. Virol.* 92, e00415–e00418. doi: 10.1128/JVI.00415-18

Canovas, A., Reverter, A., DeAtley, K. L., Ashley, R. L., Colgrave, M. L., Fortes, M. R. S., et al. (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PloS One* 9, 17. doi: 10.1371/journal.pone.0102551

Cánovas, A., Reverter, A., Kasey, L., DeAtley, K. L., Ashley, R. L., Colgrave, M. L., et al. (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PloS One* 9, e102551. doi: 10.1371/journal.pone.0102551

Castro, F. L. S., Su, S., Choi, H., Koo, E., and Kim, W. K. (2018). L-Arginine supplementation enhances growth performance, lean muscle, and bone density but not fat in broiler chickens. *Poultry Sci.* 98, 1716–1722. doi: 10.3382/ps/pey504

Chen, C., Cui, Q. M., Zhang, X., Luo, X., Liu, Y. Y., Zuo, J. B., et al. (2018a). Long non-coding RNAs regulation in adipogenesis and lipid metabolism: emerging insights in obesity. *Cell. Signalling* 51, 47–58. doi: 10.1016/j.cellsig.2018.07.012

Chen, W., Zhang, X., Li, J., Huang, S., Xiang, S., Hu, X., et al. (2018b). Comprehensive analysis of coding-lncRNA gene co-expression network uncovers conserved functional lncRNAs in zebrafish. *BMC Genomics* 19, 112. doi: 10.1186/s12864-018-4458-7

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* 132, 259–275. doi: 10.1083/jcb.132.3.259

Core Team, R. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Csorba, T., Questa, J. I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci.* 111, 16160–16165. doi: 10.1073/pnas.1419030111

Degirmenci, U., Li, J., Lim, Y. C., Siang, D. T. C., Lin, S., Liang, H., et al. (2019). Silencing an insulin-induced lncRNA, LncASIR, impairs the transcriptional response to insulin signalling in adipocytes. *Sci. Rep.* 9, 5608. doi: 10.1038/s41598-019-42162-5

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111

Doelman, J., Curtis, R. V., Carson, M., Kim, J. J. M., Metcalf, J. A., and Cant, J. P. (2015). Essential amino acid infusions stimulate mammary expression of eukaryotic initiation factor 2Bε but milk protein yield is not increased during an imbalance. *J. Dairy Sci.* 98, 4499–4508. doi: 10.3168/jds.2014-9051

Eberlein, A., Takasuga, A., Setoguchi, K., Pfuhl, R., Flisikowski, K., Fries, R., et al. (2009). Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics* 183, 951–964. doi: 10.1534/genetics.109.106476

Fleming, Ingrid. (2008). Chapter 3 "Biology of Nitric Oxide Synthases," in *Microcirculation (Second Edition)*. Eds. Tuma, R. F., Durán, W. N., and Ley, K. (San Diego, USA: Academic Press), 56–80. doi: 10.1016/B978-0-12-374530-9.00003-6

Frankish, A., Vullo, A., Zadissa, A., Yates, A., Thormann, A., Parker, A., et al. (2017). Ensembl 2018. *Nucleic Acids Res.* 46, D754–DD61. doi: 10.1093/nar/gkx1098

Haerty, W., and Ponting, C. P. (2015). Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. *RNA* 21, 333–346. doi: 10.1261/rna.047324.114

Hardie, L. C., VandeHaar, M. J., Tempelman, R. J., Weigel, K. A., Armentano, L. E., Wiggans, G. R., et al. (2017). The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. *J. Dairy Sci.* 100, 9061–9075. doi: 10.3168/jds.2017-12604

Harrell, and Frank, E. (2019). Hmisc: Harrell Miscellaneous, R-package version 4.2-0. https://cran.r-project.org/web/packages/Hmisc/index.html

Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A. et al. (2018). GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci. Rep.* 8, 14301–14301. doi: 10.1038/s41598-018-32374-6

Hou, Y. Q., Yao, K., Yin, Y. L., and Wu, G. Y. (2016). Endogenous synthesis of amino acids limits growth, lactation, and reproduction in animals. *Adv. Nutr.* 7, 331–342. doi: 10.3945/an.115.010850

Ibeagha-Awemu, E. M., Peters, S. O., Akwanji, K. A., Imumorin, I. G., and Zhao, X. (2016). High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci. Rep.* 6, 31109. doi: 10.1038/srep31109

Jin, J. J., Lv, W., Xia, P., Xu, Z. Y., Zheng, A. D., Wang, X. J., et al. (2018). Long noncoding RNA SYISL regulates myogenesis by interacting with polycomb repressive complex 2. *Proc. Natl. Acad. Sci.* 115, E9802–E9E11. doi: 10.1073/pnas.1801471115

Kashi, K., Henderson, L., Bonetti, A., Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochim. Biophys. Acta Gene Regul. Mech*. 1859, 3–15. doi: 10.1016/j.bbagrm.2015.10.010

Kenny, D. A., Fitzsimons, C., Waters, S. M., and McGee, M. (2018). Invited review: improving feed efficiency of beef cattle - the current state of the art and future challenges. *Animal* 12, 1815–1826. doi: 10.1017/S1751731118000976

Kern, C., Wang, Y., Chitwood, J., Korf, I., Delany, M., Cheng, H., et al. (2018). Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. *BMC Genomics* 19, 684. doi: 10.1186/s12864-018-5037-7

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357. doi: 10.1038/nmeth.3317

Kimball, S. R. (1999). Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.* 31, 25–29. doi: 10.1016/S1357-2725(98)00128-9

Kirchgeßner, M. (1997). Tierernährung. Frankfurt a.M., Germany: Verlags Union Agrar DLG-Verlag, 574.

Kornienko, A. E., Guenzl, P. M., Barlow, D. P., and Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* 11, 59. doi: 10.1186/1741-7007-11-59

Kramer, A., Green, J., Pollard, J. Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703

Krappmann, K., Widmann, P., Weikard, R., and Kühn, C. (2012). Variants of the bovine retinoic acid receptor-related orphan receptor C gene are in linkage disequilibrium with QTL for milk production traits on chromosome 3 in a beef × dairy crossbreed population. *Arch. Anim. Breed.* 55, 346–355. doi: 10.5194/aab-55-346-2012

Kühn, C., Bellmann, O., Voigt, J., Wegner, J., Guiard, V., and Ender, K. (2002). An experimental approach for studying the genetic and physiological background of nutrient transformation in cattle with respect to nutrient secretion and accretion type. *Arch. Anim. Breed.* 45, 14. doi: 10.5194/aab-45-317-2002

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Li, B. J., Jiang, D. L., Meng, Z. N., Zhang, Y., Zhu, Z. X., Lin, H. R., et al. (2018). Genome-wide identification and differentially expression analysis of lncRNAs in tilapia. *BMC Genomics* 19, 729. doi: 10.1186/s12864-018-5115-x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., and Shi, W. (2014). Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lind, U., Nilsson, T., McPheat, J., Stromstedt, P. E., Bamberg, K., Balendran, C., et al. (2005). Identification of the human ApoAV gene as a novel RORalpha target gene. *Biochem. Biophys. Res. Commun.* 330, 233–241. doi: 10.1016/j.bbrc.2005.02.151

Liu, P., Jin, L., Zhao, L., Long, K., Song, Y., Tang, Q., et al. (2018). Identification of a novel antisense long non-coding RNA PLA2G16-AS that regulates the expression of PLA2G16 in pigs. *Gene*. 671, 78–84. doi: 10.1016/j.gene.2018.05.114

Long, Y., Wang, X., Youmans, D. T., and Cech, T. R. (2017). How do lncRNAs regulate transcription. *Sci. Adv.* 3, eaao2110. doi: 10.1126/sciadv.aao2110

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Loy, R. E., Orynbayev, M., Xu, L., Andronache, Z., Apostol, S., Zvaritch, E., et al. (2011). Muscle weakness in Ryr1I4895T/WT knock-in mice as a result of reduced ryanodine receptor Ca2+ ion permeation and release from the sarcoplasmic reticulum. *J. Gen. Physiol.* 137, 43–57. doi: 10.1085/jgp.201010523

Marounek, M., Skřivanová, E., and Skřivanová, V. (2004). A note on the effect of caprylic acid and triacylglycerols of caprylic and capric acid on growth rate and shedding of coccidia oocysts in weaned piglets. *J. Anim. Feed Sci.* 13, 269–274. doi: 10.22358/jafs/67411/2004

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Medeiros de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., Miguel Ferreira de Camargo, G., Matos Ceacero, T., Noely dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-wide association study for carcass traits in an experimental nelore cattle population. *PloS One* 12, e0169860. doi: 10.1371/journal.pone.0169860

Miao, X., Luo, Q., Zhao, H., and Qin, X. (2016). Co-expression analysis and identification of fecundity-related long non-coding RNAs in sheep ovaries. *Sci. Rep.* 6, 39398. doi: 10.1038/srep39398

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Anderson, S. T., Islas-Trejo, A., et al. (2018). STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in brahman heifers. *Front. Genet.* 9, 87. doi: 10.3389/fgene.2018.00087

Oliveira, G. B., Regitano, L. C. A., Cesar, A. S. M., Reecy, J. M., Degaki, K. Y., Poleti, M. D., et al. (2018). Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics* 19, 126. doi: 10.1186/s12864-018-4514-3

Park, C. A., Reecy, J. M., and Hu, Z.-L. (2018). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–DD10. doi: 10.1093/nar/gky1084

Pellegrina, D.V.da Silva, Severino, P., Barbeiro, H. V., de Souza, H. P., Machado, M. C. C., Pinheiro-da-Silva, F., et al. (2017). Insights into the function of long noncoding rnas in sepsis revealed by gene co-expression network analysis. *Non-Coding RNA* 3, 5. doi: 10.3390/ncrna3010005

Percival, J. M. (2011). nNOS regulation of skeletal muscle fatigue and exercise performance. *Biophys. Rev.* 3, 209–217. doi: 10.1007/s12551-011-0060-9

Perez-Montarelo, D., Hudson, N. J., Fernandez, A. I., Ramayo-Caldas, Y., Dalrymple, B. P., and Reverter, A. (2012). Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. *PloS One* 7, e46159. doi: 10.1371/journal.pone.0046159

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Popoff, V., Adolf, F., Brugger, B., and Wieland, F. (2011). COPI budding within the Golgi stack, Cold Spring Harbor. *Perspect. Biol.* 3, a005231. doi: 10.1101/cshperspect.a005231

Raspe, E., Duez, H., Gervois, P., Fievet, C., Fruchart, J. C., Besnard, S., et al. (2001). Transcriptional regulation of apolipoprotein C-III gene expression by the orphan nuclear receptor RORalpha. *J. Biol. Chem.* 276, 2865–2871. doi: 10.1074/jbc.M004982200

Reverter, A., and Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. doi: 10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26, 896–904. doi: 10.1093/bioinformatics/btq051

Robinson, A. (2015). Quality Trim version 1.6.0. https://bitbucket.org/arobinson/qualitytrim/src/master/

Schaefer, R. J., Michno, J. M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., et al. (2018). Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30, 2922–2942. doi: 10.1105/tpc.18.00299

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18, 386–386. doi: 10.1186/s12864-017-3754-y

Serviss, J. T., Johnsson, P., and Grandér, D. (2014). An emerging role for long non-coding RNAs in cancer metastasis. *Front. Genet.* 5, 234. doi: 10.3389/fgene.2014.00234

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Stamler, J. S., and Meissner, G. (2001). Physiology of nitric oxide in skeletal muscle. *Physiol. Rev.* 81, 209–237. doi: 10.1152/physrev.2001.81.1.209

Sui, Yutong, Han, Yu, Zhao, Xingyu, Li, Dongsong, and Li, Guangyu (2019). Long non-coding RNA Irm enhances myogenic differentiation by interacting with MEF2D. *Cell Death Dis.* 10, 181–181. doi: 10.1038/s41419-019-1399-2

Sun, H. Z., Zhao, K., Zhou, M., Chen, Y., and Guan, L. L. (2019). Landscape of multi-tissue global gene expression reveals the regulatory signatures of feed efficiency in beef cattle. *Bioinformatics* 35, 1712–1719. doi: 10.1093/bioinformatics/bty883 (Oxford, England).

Sundvold, H., and Lien, S. (2001). Identification of a novel peroxisome proliferator-activated receptor (PPAR) gamma promoter in man and transactivation by the nuclear receptor RORalpha1. *Biochem. Biophys. Res. Commun.* 287, 383–390. doi: 10.1006/bbrc.2001.5602

Tang, J. E., Lysecki, P. J., Manolakos, J. J., MacDonald, M. J., Tarnopolsky, M. A., and Phillips, S. M. (2011). Bolus arginine supplementation affects neither muscle blood flow nor muscle protein synthesis in young men at rest or after resistance exercise. *J. Nutr.* 141, 195–200. doi: 10.3945/jn.110.130138

Tang, Z., Wu, Y., Yang, Y., Yang, Y.-C. T., Wang, Z., Yuan, J., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutional conservation in Sus scrofa. *Sci. Rep.* 7, 43166. doi: 10.1038/srep43166

Thornton, P. K. (2010). Livestock production: recent trends, future prospects, Philosophical Transactions of the Royal Society of London. *Ser. B Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134

Tian, W., Wang, H. R., Wu, T. Y., Ding, L. Y., Zhao, R., Khas, E., et al. (2017). Milk protein responses to balanced amino acid and removal of Leucine and Arginine supplied from jugular-infused amino acid mixture in lactating dairy cows. *J. Anim. Physiol. Anim. Nutr.* 101, e278–ee87. doi: 10.1111/jpn.12603

Tunnacliffe, A., Pensotti, V., and Radice, P. (1996). The coatomer protein delta-COP, encoded by the archain gene, is conserved across diverse eukaryotes. *Mamm. Genome* 7, 784–786. doi: 10.1007/s003359900234

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell.* 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., and Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550. doi: 10.1016/j.cell.2011.11.055

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings Bioinf.* 19, 575–592. doi: 10.1093/bib/bbw139

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. doi: 10.1038/nrg2538

Vu-Dac, N., Gervois, P., Grotzinger, T., De Vos, P., Schoonjans, K., Fruchart, J. C., et al. (1997). Transcriptional regulation of apolipoprotein A-I gene expression by the nuclear receptor RORalpha. *J. Biol. Chem.* 272, 22401–22404. doi: 10.1074/jbc.272.36.22401

Walsh, D., and Mohr, I. (2014). Coupling 40S ribosome recruitment to modification of a cap-binding initiation factor by eIF3 subunit e. *Genes Dev.* 28, 835–840. doi: 10.1101/gad.236752.113

Wang, C.-H., Shi, H.-H., Chen, L.-H., Li, X.-L., Cao, G.-L., and Hu, X.-F. (2019). Identification of key lncRNAs associated with atherosclerosis progression based on public datasets. *Front. Genet.* 10, 123. doi: 10.3389/fgene.2019.00123

Weikard, R., Altmaier, E., Suhre, K., Weinberger, K. M., Hammon, H. M., Albrecht, E., et al. (2010). Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. *Physiol. Genomics* 42a, 79–88. doi: 10.1152/physiolgenomics.00120.2010

Weikard, R., Goldammer, T., Brunner, R. M., and Kuehn, C. (2012). Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiol. Genomics* 44, 728–739. doi: 10.1152/physiolgenomics.00007.2012

Weikard, R., Hadlich, F., Hammon, H. M., Frieten, D., Gerbert, C., Koch, C., et al. (2018). Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. *Oncotarget* 9, 21052–21069. doi: 10.18632/oncotarget.24898

Widmann, P., Nuernberg, K., Kuehn, C., and Weikard, R. (2011). Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genet.* 12, 13. doi: 10.1186/1471-2156-12-96

Widmann, P., Reverter, A., Fortes, M. R. S., Weikard, R., Suhre, K., Hammon, H., et al. (2013). A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics* 14, 798. doi: 10.1186/1471-2164-14-798

Widmann, P., Reverter, A., Weikard, R., Suhre, K., Hammon, H. M., Albrecht, E., et al. (2015). Systems biology analysis merging phenotype, metabolomic and genomic data identifies non-smc condensin i complex, subunit G (NCAPG) and cellular maintenance processes as major contributors to genetic variability in bovine feed efficiency. *PloS One* 10, 22. doi: 10.1371/journal.pone.0124574

Wu, G., Bazer, F. W., Davis, T. A., Kim, S. W., Li, P., Marc Rhoads, J., et al. (2009). Arginine metabolism and nutrition in growth, health and disease. *Amino Acids* 37, 153–168. doi: 10.1007/s00726-008-0210-y

Wucher, V., Legeai, F., Hedan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45, e57. doi: 10.1093/nar/gkw1306

Yang, L., Li, P., Yang, W., Ruan, X., Kiesewetter, K., Zhu, J., et al. (2016). Integrative transcriptome analyses of metabolic responses in mice define pivotal LncRNA metabolic regulators. *Cell Metab.* 24, 627–639. doi: 10.1016/j.cmet.2016.08.019

Zeng, Y., Ren, K., Zhu, X., Zheng, Z., and Yi, G. (2018). "Chapter one long noncoding RNAs," in *Advances in Lipid Metabolism*. Ed. Gregory, S. (Makowski: Advances in Clinical Chemistry (Elsevier)). doi: 10.1016/bs.acc.2018.07.001

Zhang, Y., Papazyan, R., Damle, M., Fang, B., Jager, J., Feng, D., et al. (2017). The hepatic circadian clock fine-tunes the lipogenic response to feeding through RORalpha/gamma. *Genes Dev.* 31, 1202–1211. doi: 10.1101/gad.302323.117

Zhu, M., Liu, J. F., Xiao, J., Yang, L., Cai, M. X., Shen, H. Y., et al. (2017). Lnc-mg is a long non-coding RNA that promotes myogenesis. *Nat. Commun.* 8, 11. doi: 10.1038/ncomms14718

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nolte, Weikard, Brunner, Albrecht, Hammon, Reverter and Kühn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 November 2019
doi: 10.3389/fgene.2019.01215

Expression Quantitative Trait Loci in Equine Skeletal Muscle Reveals Heritable Variation in Metabolism and the Training Responsive Transcriptome

*Gabriella Farries[1], Kenneth Bryan[1], Charlotte L. McGivney[2], Paul A. McGettigan[1], Katie F. Gough[1], John A. Browne[1], David E. MacHugh[1,3], Lisa Michelle Katz[2] and Emmeline W. Hill[1,4]**

[1] UCD School of Agriculture and Food Science, University College Dublin, Dublin, Ireland, [2] UCD School of Veterinary Medicine, University College Dublin, Dublin, Ireland, [3] UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland, [4] Research and Development, Plusvital Ltd., Dublin, Ireland

OPEN ACCESS

Edited by:
Fabyano Fonseca Silva, Universidade Federal de Viçosa, Brazil

Reviewed by:
Eric Barrey, INRA UMR1313 Genetique Animale et Biologie Integrative, France
Yame Fabres Robaina Sancler-Silva, Universidade Federal de Viçosa, Brazil

***Correspondence:**
Emmeline W. Hill
emmeline.hill@ucd.ie

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 09 July 2019
Accepted: 04 November 2019
Published: 26 November 2019

Citation:
Farries G, Bryan K, McGivney CL, McGettigan PA, Gough KF, Browne JA, MacHugh DE, Katz LM and Hill EW (2019) Expression Quantitative Trait Loci in Equine Skeletal Muscle Reveals Heritable Variation in Metabolism and the Training Responsive Transcriptome. Front. Genet. 10:1215. doi: 10.3389/fgene.2019.01215

While over ten thousand genetic loci have been associated with phenotypic traits and inherited diseases in genome-wide association studies, in most cases only a relatively small proportion of the trait heritability is explained and biological mechanisms underpinning these traits have not been clearly identified. Expression quantitative trait loci (eQTL) are subsets of genomic loci shown experimentally to influence gene expression. Since gene expression is one of the primary determinants of phenotype, the identification of eQTL may reveal biologically relevant loci and provide functional links between genomic variants, gene expression and ultimately phenotype. Skeletal muscle (gluteus medius) gene expression was quantified by RNA-seq for 111 Thoroughbreds (47 male, 64 female) in race training at a single training establishment sampled at two time-points: at rest (n = 92) and four hours after high-intensity exercise (n = 77); n = 60 were sampled at both time points. Genotypes were generated from the Illumina Equine SNP70 BeadChip. Applying a False Discovery Rate (FDR) corrected P-value threshold ($P_{FDR} < 0.05$), association tests identified 3,583 *cis*-eQTL associated with expression of 1,456 genes at rest; 4,992 *cis*-eQTL associated with the expression of 1,922 genes post-exercise; 1,703 *trans*-eQTL associated with 563 genes at rest; and 1,219 *trans*-eQTL associated with 425 genes post-exercise. The gene with the highest *cis*-eQTL association at both time-points was the endosome-associated-trafficking regulator 1 gene (*ENTR1*; Rest: $P_{FDR} = 3.81 \times 10^{-27}$, Post-exercise: $P_{FDR} = 1.66 \times 10^{-24}$), which has a potential role in the transcriptional regulation of the solute carrier family 2 member 1 glucose transporter protein (SLC2A1). Functional analysis of genes with significant eQTL revealed significant enrichment for cofactor metabolic processes. These results suggest heritable variation in genomic elements such as regulatory sequences (e.g. gene promoters, enhancers, silencers), microRNA and transcription factor genes, which are associated with metabolic function and may have roles in determining end-point muscle and athletic performance phenotypes in Thoroughbred horses. The incorporation of the eQTL identified with

genome and transcriptome-wide association may reveal useful biological links between genetic variants and their impact on traits of interest, such as elite racing performance and adaptation to training.

Keywords: expression quantitative trait loci, gene expression, RNA sequencing, horse, exercise, aerobic metabolism

INTRODUCTION

In the 6,000 years since horses were first domesticated on the Eurasian steppe, there has been strong artificial selection for various athletic traits (Levine, 1999). Selection for athleticism is perhaps most clearly manifested in the Thoroughbred, which has undergone over 300 years of intense selection for speed and racing performance (Willett, 1975; Todd et al., 2018). As a result the Thoroughbred has a highly developed musculature, with a skeletal muscle mass ~10% greater than other horse breeds (~55% compared to ~42%) (Gunn, 1987), accompanied by decreased body fat (Kearns et al., 2002), superior glycogen storage capacity (Votion et al., 2012), increased mitochondrial volume (compared to other mammals) (Kayar et al., 1989) and a high degree of plasticity in skeletal muscle fibre composition (Rivero, 2004).

The response of equine skeletal muscle to training has been well studied (Snow et al., 1985; Snow, 1994). These responses in general increase the oxidative capacity of the muscle, such as fibre type switching from fast-twitch glycolytic fibres to slow-twitch, high-oxidative fibres (Snow, 1994; Serrano et al., 2000), an increase in oxidative phosphorylation (Snow et al., 1985; Votion et al., 2012) and increased mitochondrial volume (Tyler et al., 1998). Training also elicits an increase in skeletal muscle mass (Rivero et al., 1996), mediated through hyperplastic growth as opposed to marked hypertrophy (Rivero et al., 1996; Rivero et al., 2002).

The transcriptional response to exercise and training in skeletal muscle has been studied in the Thoroughbred (McGivney et al., 2009; Eivers et al., 2010; McGivney et al., 2010; Eivers et al., 2012; Bryan et al., 2017). Initially, reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) was used to quantify expression of 18 candidate genes in response to a standardised exercise test on a high-speed treadmill (Eivers et al., 2010). Significant differential expression of creatine kinase M-type (*CKM*), cytochrome c oxidase subunit 4I1 (*COX4I1*), cytochrome c subunit 4I2 (*COX4I2*), pyruvate dehydrogenase kinase 4 (*PDK4*), PPARG coactivator 1 alpha (*PPARG1A*) and solute carrier family 24 member 4 (*SLC2A4*) four hours post-exercise was detected. *PPARG1A* is a transcription factor downstream of hypoxia-inducible factor (HIF), activation of *PPARG1A via* HIF in response to exercise induces downstream adaptations in oxidative phosphorylation (Arany, 2008). The differentially expressed genes were downstream targets of HIF or related to oxidative phosphorylation or muscle substrate use (Kraniou et al., 2006) The availability of a dedicated equine microarray allowed gene expression to be measured across 9,333 expressed sequence tags (ESTs). This technology was then used to examine the changes in gene expression induced by exercise, without *a priori* knowledge of the genes involved (McGivney et al., 2009). Analysis of the differentially expressed genes showed a functional enrichment of genes involved in insulin signalling, focal adhesion, hypertrophic and apoptotic pathways. Digital gene expression was used to investigate the transcriptional response to a ten-month training protocol (McGivney et al., 2010), identifying functional enrichment of genes relevant to aerobic metabolism. More recently, RNA sequencing (RNA-seq) was used to investigate the response to both exercise and training, and a network biology approach was employed to identify relevant functional modules that highlighted the role of autophagy (Bryan et al., 2017) While these studies provide insight to the genes involved in the transcriptional response to exercise, they do not reveal whether there is variation in the transcriptional response among individuals and how this may influence skeletal muscle function.

Expression quantitative trait loci (eQTL) are genomic variants, typically single nucleotide polymorphism (SNPs), that are associated with variation in RNA transcript abundance. Jansen and Nap (2001) introduced the concept of 'genetical genomics' where genomic loci were associated with cellular intermediates, such as transcript abundance, to catalogue functional relevance for non-coding variants. These measurements at a cellular level then act as endophenotypes, which are heritable, intermediate phenotypes This was a particularly important development because the clear majority (>85%) of QTLs detected in genome-wide association studies (GWAS) are located in non-coding regions (Hindorff et al., 2009; Brown et al., 2013).

Cis-eQTL are genetic variants that alter gene expression in an allele-specific manner and are typically located in gene regulatory regions (Wittkopp, 2005; Westra and Franke, 2014). Identification of true *cis*-eQTL requires aligning reads to their chromosome of origin; consequently, many studies have by convention defined any eQTLs within 1 Mb of the transcription start site (TSS) of the gene they act on as *cis*. Conversely, *trans*-eQTL act in a less direct manner, altering the expression of a secondary genome product—for example, a transcription factor or a microRNA—that regulates expression of a distant gene elsewhere in the genome (Wittkopp, 2005).

The study of eQTL in skeletal muscle to-date has been largely to investigate functional variants in the pathogenesis of type II diabetes (T2D) in humans (Mason et al., 2011; Keildson et al., 2014b; Sajuthi et al., 2016). While GWAS for T2D have identified loci associated with disease risk, these studies have not provided information on the function of these variants or the mechanism by which they contribute to disease. Keildson et al. (2014b) performed an eQTL investigation using skeletal muscle biopsies from 104 human subjects and identified an association between the rs4547172 SNP and muscle phosphofructokinase gene (*PFKM*) expression. Furthermore, the study found that

increased expression of *PFKM* was associated with increased resting plasma insulin (an endophenotype) and T2D (an end-point phenotype). This example shows that an eQTL approach can identify functional links between genomic variants, gene expression, endophenotypes, and ultimately, disease.

Variation in human gene expression has been found to be highly heritable (Monks et al., 2004; Stranger et al., 2007; Wright et al., 2014). Given the influence of gene expression on phenotype, detection of heritable variation in skeletal muscle gene expression may provide insight into genomic loci contributing to variation in exercise and performance related phenotypes.

In this study, we hypothesised that there is heritable variation in the Thoroughbred skeletal muscle transcriptional response to exercise and training, and that this variation may have implications for athletic performance.

METHODS

Ethics Statement

University College Dublin Animal Research Ethics Committee approval (AREC-P-12-55-Hill), a licence from the Department of Health (B100/3525), and informed owner consent were obtained.

Cohort

Skeletal muscle biopsy samples (gluteus medius) were collected from 111 horses (47 male, 64 female) born between 2011 and 2012. All horses were based at a single training yard, under the supervision of a single trainer and under similar management and feeding regimes. The 111 horses used for the study were produced from 19 different sires and 94 different dams.

Biopsies were collected at two time points: untrained at rest (UR) and untrained four hours post-exercise (UE). Of the 111 horses, 60 were sampled at both time points. In total 92 UR samples and 77 UE samples were collected. The horses were defined as untrained because they had completed ≤ four sprint exercise bouts (e.g., work days) prior to sampling. The number of prior work days and days of submaximal prior training prior to sampling were recorded. Horses were defined as untrained in order to integrate results with those of Bryan et al. (2017), where the untrained cohort had performed only 1–2 work days prior to sampling, and the trained cohort had completed a mean of 15.1 work days prior to sampling (SD = 9.1).

Exercise Test

The exercise stimulus was an intense sprint bout of exercise (work day) undertaken as part of normal training. The training regime for horses is submaximal training at canter six times per week, with work days being introduced and replacing one to two submaximal bouts per week. On a work day horses were initially walked on an automated horse walker for 30–60 min, followed by 5–10 min of walking in hand. Under saddle there was an initial warm-up period of 300 m walk and 700 m of trot and slow canter down the incline of the track. The work day was performed on a 1,500 m all-weather woodchip gallop track, with the final 800 m straight set on a 2.7% incline. The sprint portion of the exercise bout consisted of the horses galloping at high intensity for 800-1,000 m up the incline of the gallop. In a larger cohort of horses (n = 294) from the same training establishment, the work day was characterised using concurrent global positioning system (GPS) and heart rate monitoring (Farries et al., 2019). From 2,900 GPS recordings the mean peak speed was 16.36 m/s (range: 14.23–17.63 m/s). Of these 2,900 recordings 1,056 had simultaneous heart rate recordings, with a mean peak heart rate of 219 beats per minute (range: 182–237).

For 34/77 UE horses, whole blood was collected at rest and five minutes post-exercise into fluoride oxalate tubes. Samples were centrifuged, and plasma lactate concentrations measured on-site using a YSI2300 STAT PLUS auto analyser (YSI UK Ltd, Hampshire, UK). These measurements were used to validate the intensity of the exercise test performed.

Biopsy Sampling

Percutaneous needle muscle biopsies (approximately 300 mg) were obtained from the ventral compartment of the middle gluteal muscle using the method described by Valette et al. (1999). All UR samples were collected between 7:30 am and 11:30 am. UE samples were taken four hours after completion of the exercise test, as this has previously been shown to be a timepoint where the greatest change in gene expression in response to acute exercise was observed (McGivney et al., 2009; Eivers et al., 2010). Muscle samples were stored in RNAlater (Thermo Fisher, Massachusetts, USA) for 24 hours at 4°C then stored at −20°C prior to RNA extraction.

RNA Extraction and Quality Control

Total RNA was extracted from approximately 70 mg tissue using a protocol combining TRIzol reagent (Thermo Fisher), DNase I treatment (Qiagen, Hilden, Germany) and an RNeasy Mini-Kit (Qiagen). RNA was quantified using a Nano Drop ND1000 spectrophotometer V 3.5.27 (Thermo Fisher). RNA quality was assessed using the RNA integrity number (RIN) on an Agilent Bioanalyser with the RNA 6000 Nano LabChip kit6 (Agilent, Cork, Ireland).

RNA Sequencing

Indexed, strand-specific Illumina sequencing libraries were prepared using the TruSeq Stranded mRNA Library Preparation Kit LT (Illumina, San Diego, CA, USA). Libraries were pooled with 18–20 indexed libraries per pool and sequenced on an Illumina HiSeq 2500 using a Rapid Run flow cell and reagents to generate 100 bp paired-end reads. Each pool was sequenced across both lanes of the flow cell (dual lane loading). Demultiplexed sequence data was then converted to FASTQ format. Sequencing was performed by the Research Technology Support Facility, Michigan State University.

RNA-Seq Data Workflow

Quality control of the sequence reads was performed using *FastQC* [version: 0.11.5] (Andrews, 2010). *STAR* aligner [version: 2.5.2b] (Dobin et al., 2013) was used to map reads to the Equine reference genome EquCab2 (Ensembl release 62).

After mapping, *featureCounts* [version: 1.5.0] was used to assign reads to genes (Liao et al., 2014). Data for each sample from each sequencing lane was then merged where concordance was >99% between lanes. Count data was parsed using a custom script, then small non-coding RNA were filtered using *BiomaRt* (Durinck et al., 2009). Assessment of the count data and multidimensional scaling were performed using *edgeR* (Robinson et al., 2010). Results of the multidimensional scaling were visualised using *ggplot2* (Wickham, 2009). Count data was quantile normalised using *preprocessCore* [version: 1.40.0] (Bolstad, 2017) within the R environment [version: 3.5.1] (R Core Team, 2017), and the $\log_2$ of quantile-normalised count data calculated.

Genotyping

Genomic DNA was extracted from whole blood using the Maxwell 16 automated DNA purification system (Promega, Madison, WI, USA). Horses were genotyped on the Illumina Equine SNP70 BeadChip (Illumina). A genetic versus phenotypic sex check was performed. SNPs with a genotyping rate of <95%, and individuals with a genotyping rate <95% were excluded. SNPs with a minor allele frequency (MAF) < 0.10 were removed. Using these quality-controlled SNPs, identity by state (IBS) distances between individuals were calculated using the 'genome' function in *PLINK* [version 1.09] (Chang et al., 2015). The remaining 43,988 SNPs were then pruned based on pairwise linkage disequilibrium (LD) using a sliding window with an LD threshold of $r^2 > 0.7$, a window size of 50, and a step of 5 in *PLINK*. A set of 15,995 SNPs were used for the eQTL analysis. Pruning was undertaken due to the large spanning of LD within the Thoroughbred, with previous work validating the use of <15,000 SNPs to capture the majority of genetic variation (Corbin et al., 2014; Schaefer et al., 2017).

eQTL Analysis

eQTL were determined using a linear model within *matrixEQTL* [version: 2.1.1] (Shabalin, 2012); including sex and age at sampling (days) as covariates. As samples had been included in two separate sample pools which were sequenced separately, the sequencing batch for each sample was also included as a covariate. Tests of association were corrected using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) and eQTL with a corrected *P*-value (P_{FDR}) < 0.05 were catalogued for UR and UE samples separately. eQTL located within 1 Mb of the transcription start site (TSS) of the gene they were associated were designated as *cis*, and those located >1Mb from the TSS were designated *trans*, in keeping with human eQTL studies (Lonsdale et al., 2013). Significant results were then compared against genes previously identified in the skeletal muscle transcriptional response to acute, high-intensity exercise (a work day; 3,241 genes) and transcriptional response to a six-month period of training (3,405 genes) (Bryan et al., 2017).

Functional Enrichment Analysis

Genes with significant eQTL were investigated for enrichment of biological processes using gene ontology (GO) categories (Ashburner et al., 2000) with the *clusterProfiler* package [version: 3.10.1] (Yu et al., 2012) within the R environment. Equine Ensembl IDs were mapped to annotated human orthologs, retrieved from the BioMart database (Kasprzyk, 2011) and GO enrichment performed using the annotation from the human genome annotation package *org.Hs.eg.db* [version: 2.12.0] (Carlson, 2019). The background gene set was the complement of genes expressed in skeletal muscle identified in this study (13,384 genes; 12,707 mapped to human orthologs). A threshold for significant enrichment was set at <0.05 after adjustment using the Benjamini-Hochberg procedure (P_{FDR}) (Benjamini and Hochberg, 1995). The number of genes assigned to each Biological Process (Gene count) and proportion of genes associated with that cluster out of all the genes expressed (Gene ratio) were also reported. Results were visualised using the *clusterProfiler* package (Yu et al., 2012).

RESULTS

Cohort

UR horses had a mean age of 611.7 days (range: 513–787 days), UE horses had a mean age of 757.5 days (range: 617–1,283 days). Dates of commencing preparatory training were available for 90 of the UR horses; 21 of the UR horses were sampled prior to breaking, 69 were sampled after breaking with a mean of 41.5 days after commencing preparatory training (range: 5-154 days) (**Table 1**). UE horses were sampled on average 156.6 days after commencing preparatory training (range: 31–307). UR horses had an average of 41.5 days submaximal training (range: 5–154) and UE had on average 48.6 (range: 19–152) (**Table 1**). UR horses had completed a mean of 0.3 work days (range: 0–4), UE horses completed a mean of 0.5 WDs prior to sampling (range: 0–3) (**Table 2**). A subset of 34 of the UE horses had a mean peak post-exercise plasma lactate concentration of 28.2 mmol/L, and a mean resting plasma concentration of 0.4 mmol/L. All RNA samples used for RNA-seq had a RIN greater than 7.0, the UR cohort had a mean RIN of 8.0 (range: 7.2–9.3) and the UE cohort had a mean RIN of 8.1 (range: 7.0–9.3). Multi-dimensional scaling was used to visually inspect the count data, showing separation of untrained resting and untrained post-exercise samples (**Figure S1**).

Analysis of the genetic relatedness of the cohort showed the mean IBS distance between individuals was 0.69 and ranged from 0.64–0.85 (SD = 0.03). Of the 19 sires represented in the cohort, the top six sires in terms of number of progent represented had 39, 23, 14, 10, 6, and 4 progeny. There were two sires with two progeny and the remaining ten sires had one offspring each. There were 12 full siblings in the cohort and 34 half-siblings by dam.

eQTL Discovery

Using the full complement of 13,384 genes, 3,582 *cis*-eQTL and 1,703 *trans*-eQTL were detected in UR samples ($P_{FDR} < 0.05$). The 3,582 *cis*-eQTL were associated with expression of 1,456 genes. The gene with the strongest *cis*-eQTL (BIEC2-707785) in UR horses was the endosome associated trafficking regulator 1 gene (*ENTR1*; $P_{FDR} = 3.81 \times 10^{-27}$) (**Figure S2**, **Table 3**). GO enrichment analysis of the *cis* regulated genes in UR samples showed that the most significantly enriched Biological Process was 'GO:0006805 xenobiotic metabolic process ' ($P_{FDR} = 3.02 \times 10^{-7}$, Gene Ratio =

TABLE 1 | Description of prior training completed by horses in the untrained resting and untrained post-exercise cohorts.

	Preparatory training		Submaximal training	
Untrained resting	No. horses sampled before entering preparatory training	21	No. horses sampled before entering submaximal training	52
	No. horses sampled after entering preparatory training	69	No. horses sampled after entering submaximal training	16
	Days preparatory training prior to sampling	Mean: 31.82 SD: 40.78 Range: 5-154	Days submaximal training prior to sampling	Mean: 11.44 SD: 26.16 Range: 19-152
Untrained post-exercise	No. horses sampled before entering preparatory training	0	No. horses sampled before entering submaximal training	0
	No. horses sampled after entering preparatory training	74	No. horses sampled after entering submaximal training	76
	Days preparatory training prior to sampling	Mean: 156.61 SD: 63.91 Range: 31-307	Days submaximal training prior to sampling	Mean: 124.59 SD: 77.61 Range: 7-302

TABLE 2 | Number of prior high-intensity sprint bouts (work days, WDs) completed prior to sampling for horses within untrained resting and untrained post-exercise cohorts.

Untrained resting		Untrained post-exercise	
Prior WDs	**No. horses**	**Prior WDs**	**No. horses**
0	78	0	47
1	7	1	24
2	5	2	4
3	2	3	1
4	1	4	0

33/1,614). 'GO:0051186 cofactor metabolic process' (P_{FDR} = 1.42 × 10^{-4}) was also significantly enriched and had the largest Gene Ratio (105/1,614) (**Figure 1**, **Table S1**).

In the UR cohort 1,219 *trans*-eQTL were associated with 425 genes. The majority 70.39% (858) were located on the same chromosome as the associated gene, and 29.61% (361) were associated with genes located on different chromosomes. The most significant *trans*-eQTL was BIEC2-526896 on ECA20 and expression of the DEAH-box helicase 16 gene (*DHX16*) also located on ECA20 1.49 Mb downstream from BIEC2-526896 (P_{FDR} = 3.50 × 10^{-17}) (**Table 4**). Functional analysis of the *trans* eGenes showed enrichment of 'interferon-gamma-mediated signalling' (P_{FDR} = 6.06 × 10^{-4}, Gene Ratio = 13/340) (**Figure 2**, **Table S2**). The functional categories with the highest Gene Ratios were 'cofactor metabolic process' (P_{FDR} = 0.01, Gene Ratio = 29/340) and 'monocarboxylic acid metabolic process' (P_{FDR} = 0.02, Gene Ratio = 29/340) (**Figure 2**, **Table S2**).

In the UE cohort 4,992 *cis*-eQTL were associated with the expression of 1,922 genes. The most significant *cis*-eQTL was BIEC2-707785 on ECA25 and *ENTR1* (P_{FDR} = 1.66 × 10^{-24}) (**Figure S3**), as was the case in the UR cohort (**Table 3**). The strongest *trans*-eQTL association was between BIEC2-165011

TABLE 3 | Top 10 genes by strongest *cis*-eQTL association.

SNP	Ensembl Gene ID	Gene	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting										
BIEC2-707785	ENSECAG00000010463	*ENTR1*	25	37504251	3.81E-27	25	37555303	37560097	1	-51052
BIEC2-1078267	ENSECAG00000012661	NA	8	1998191	1.46E-24	8	1610405	1612946	-1	387786
BIEC2-1078267	ENSECAG00000021215	NA	8	1998191	3.61E-24	8	2897771	2900415	-1	-899580
BIEC2-1078267	ENSECAG00000020446	NA	8	1998191	5.51E-23	8	2440998	2443623	-1	-442807
BIEC2-1078267	ENSECAG00000019976	NA	8	1998191	3.23E-22	8	2341391	2343774	-1	-343200
BIEC2-737866	ENSECAG00000024388	*BPIFC*	28	30492661	3.20E-21	28	30663308	30696619	-1	-170647
BIEC2-1078267	ENSECAG00000014546	NA	8	1998191	5.14E-21	8	1868952	1871600	-1	129239
BIEC2-235541	ENSECAG00000019034	*NMRAL1*	13	38538682	2.28E-18	13	38492654	38500913	1	46028
BIEC2-312869	ENSECAG00000021118	NA	15	28568686	5.02E-18	15	29280622	29284093	-1	-711936
BIEC2-163619	ENSECAG00000003428	NA	11	58806041	8.51E-18	11	57863227	57863916	-1	942814
Untrained post-exercise										
BIEC2-707785	ENSECAG00000010463	*ENTR1*	25	37504251	1.66E-24	25	37555303	37560097	1	-51052
BIEC2-737866	ENSECAG00000024388	*BPIFC*	28	30492661	2.41E-21	28	30663308	30696619	-1	-170647
BIEC2-240006	ENSECAG00000017792	*PARN*	13	30544927	2.41E-21	13	30444515	30591937	1	100412
BIEC2-1078267	ENSECAG00000012661	NA	8	1998191	4.18E-18	8	1610405	1612946	-1	387786
BIEC2-741205	ENSECAG00000021492	*ADSL*	28	36754782	6.53E-18	28	37011649	37027102	1	-256867
BIEC2-328876	ENSECAG00000020459	*SEC13*	16	6720019	7.95E-18	16	6667915	6693504	1	52104
TBIEC2-918150	ENSECAG00000000861	*CHCHD3*	4	87261712	7.95E-18	4	86938001	87207374	-1	323711
BIEC2-166011	ENSECAG00000016949	NA	11	49589952	7.95E-18	11	49427265	49438385	1	162687
BIEC2-898764	ENSECAG00000013187	*LPGAT1*	5	26823295	3.92E-17	5	26734686	26804173	1	88609
BIEC2-62887	ENSECAG00000016304	*CASC4*	1	144947609	4.14E-17	1	1.45E+08	144833523	-1	220296

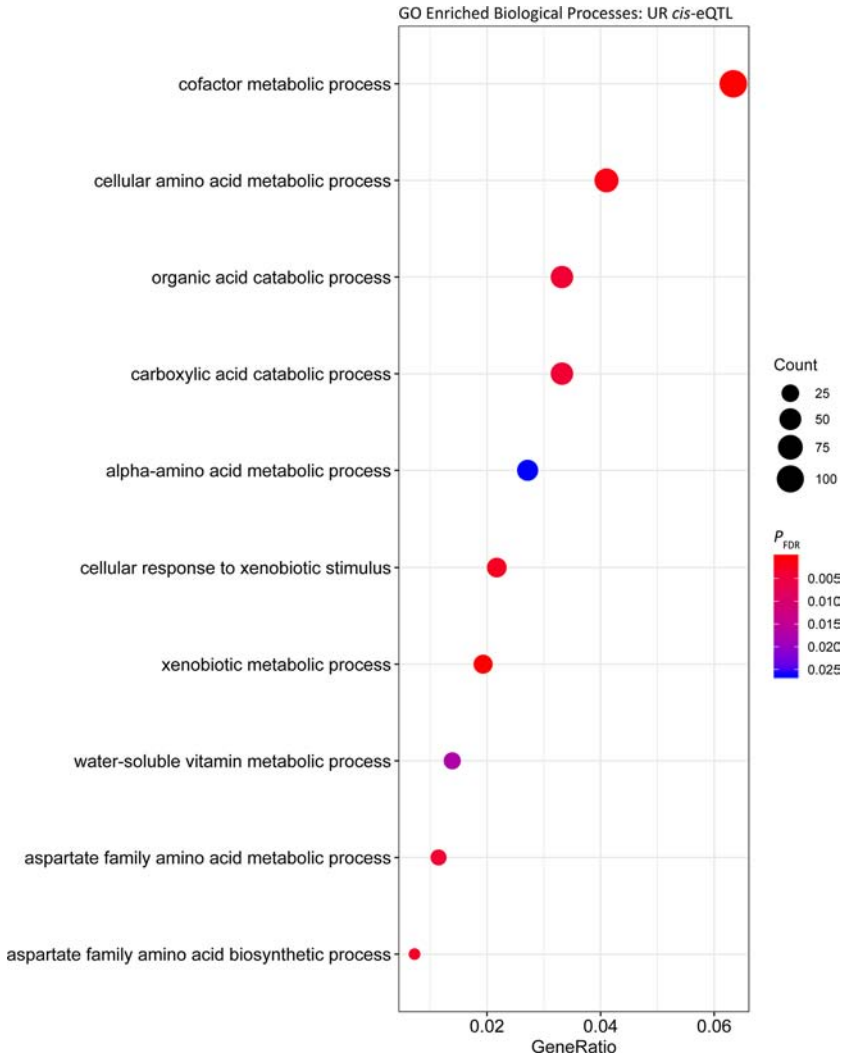


FIGURE 1 | Gene ontology enrichment of biological processes for genes under *cis* regulation in the untrained resting cohort.

on ECA11 and transcript ENSECAG00000016949 ($P_{FDR} = 1.12 \times 10^{-13}$) (**Table 4**). Similar to UR samples, the majority (75.45%; 544) of *trans*-eQTL were located on the same chromosome and 24.55% (177) were on different chromosomes.

Analysis of the *cis* regulated genes in UE samples showed that similar to the UR cohort, the most significantly enriched Biological Process was 'cofactor metabolic process' ($P_{FDR} = 6.40 \times 10^{-7}$, Gene Ratio = 112/1,579) (**Figure 3**, **Table S3**). Comparable results were obtained for enrichment of Biological Processes among putative *trans* regulated genes, with 'cofactor metabolic process' the most significantly enriched (**Figure 4**; $P_{FDR} = 5.06 \times 10^{-7}$, Gene Ratio = 33/235).

Genetic Regulation of Exercise Relevant Genes

The total set of genes with expression changes associated with eQTLs (i.e. eGenes) were queried against genes that we have reported from the same dataset to be differentially expressed post-exercise in a sample of 39 Thoroughbreds (Bryan et al., 2017). Of the 3,582 UR *cis*-eQTL, 913 were associated with genes differentially expressed in response to exercise. The most significant association was between BIEC2-285235 and the CCR4-NOT transcription complex subunit 11 gene (*CNOT11*; $P_{FDR} = 3.00 \times 10^{-15}$) (**Table 5**). Of the 1,703 UR *trans*-eQTL, 144 were associated with exercise relevant genes. The most significant *trans*-eQTL was between BIEC2-1061469 and the TAL bHLH transcription factor 2 gene (*TAL2*; $P_{FDR} = 3.03 \times 10^{-10}$) (**Table 6**).

Within the UE cohort 4,992 *cis*-eQTL were identified, 1,132 of which were associated with genes differentially expressed post-exercise. The strongest association was between BIEC2-240006 and the polyA-specific ribonuclease gene (*PARN*; $P_{FDR} = 2.41 \times 10^{-21}$) (**Table 5**). Of the UR trans-eQTL, 121 eQTL were associated with eGenes in the transcriptional exercise response. The strongest *trans*-eQTL in the UE cohort was BIEC2-1053404, associated with expression of the peroxiredoxin 2 gene (*PRDX2*; $P_{FDR} = 1.51 \times 10^{-8}$) (**Table 6**).

TABLE 4 | top ten genes by strongest *trans*-eQTL association.

SNP	Ensembl Gene ID	Gene	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting										
BIEC2-526896	ENSECAG00000008537	*DHX16*	20	28206968	3.50E-17	20	29693915	29708818	-1	-1486947
BIEC2-218753	ENSECAG00000015585	*TBL2*	13	9657509	8.69E-14	13	11241845	11246395	-1	-1584336
BIEC2-526896	ENSECAG00000015782	NA	20	28206968	5.20E-13	20	29364710	29368260	1	-1157742
BIEC2-526896	ENSECAG00000015505	NA	20	28206968	3.57E-12	20	30173637	30178981	-1	-1966669
BIEC2-554291	ENSECAG00000019318	NA	20	29116874	3.57E-12	20	32450503	32452484	-1	-3333629
BIEC2-991035	ENSECAG00000012956	*DCPS*	7	31604226	9.10E-12	7	35117723	35153555	1	-3513497
BIEC2-526896	ENSECAG00000021750	NA	20	28206968	2.27E-11	20	31193941	31197235	-1	-2986973
BIEC2-526791	ENSECAG00000009368	NA	20	27782384	2.27E-11	20	29049395	29052907	1	-1267011
BIEC2-554291	ENSECAG00000021750	NA	20	29116874	4.38E-11	20	31193941	31197235	-1	-2077067
BIEC2-31048	ENSECAG00000024546	NA	1	73056328	5.71E-11	1	70638960	70656636	-1	2417368
Untrained post-exercise										
UKUL2765	ENSECAG00000010213	*MCCC2*	14	93893549	1.12E-13	14	92304084	92460487	-1	1589465
BIEC2-3702	ENSECAG00000004989	NA	1	9825107	4.38E-12	1	96257111	96258450	1	-86432004
BIEC2-991035	ENSECAG00000012956	*DCPS*	7	31604226	2.03E-11	7	35117723	35153555	1	-3513497
BIEC2-526896	ENSECAG00000015505	NA	20	28206968	6.37E-11	20	30173637	30178981	-1	-1966669
BIEC2-210079	ENSECAG00000012497	*LIMK1*	13	10604233	6.76E-11	13	11610902	11634978	1	-1006669
BIEC2-554291	ENSECAG00000019318	NA	20	29116874	2.91E-10	20	32450503	32452484	-1	-3333629
BIEC2-526896	ENSECAG00000015782	NA	20	28206968	4.06E-10	20	29364710	29368260	1	-1157742
BIEC2-554291	ENSECAG00000015505	NA	20	29116874	8.06E-10	20	30173637	30178981	-1	-1056763
BIEC2-63245	ENSECAG00000016304	*CASC4*	1	1.48E+08	1.04E-09	1	1.45E+08	1.45E+08	-1	3045045
BIEC2-3702	ENSECAG00000009972	*LHPP*	1	9825107	2.16E-09	1	8372229	8511110	-1	1452878

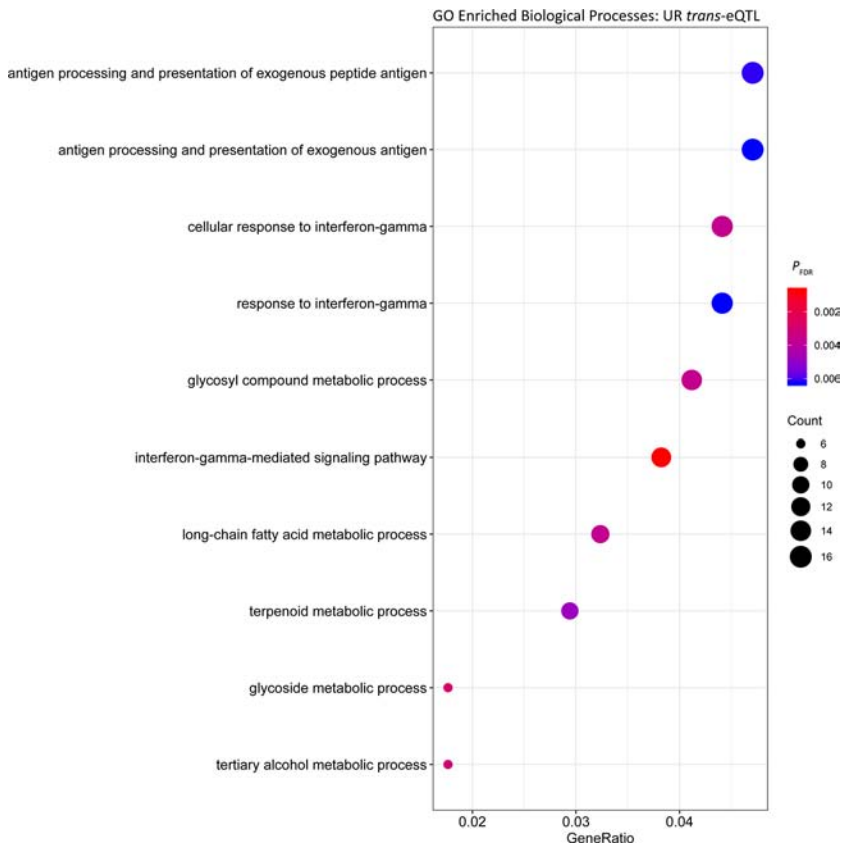


FIGURE 2 | Gene ontology enrichment of biological processes for genes under *trans* regulation in the untrained resting cohort.

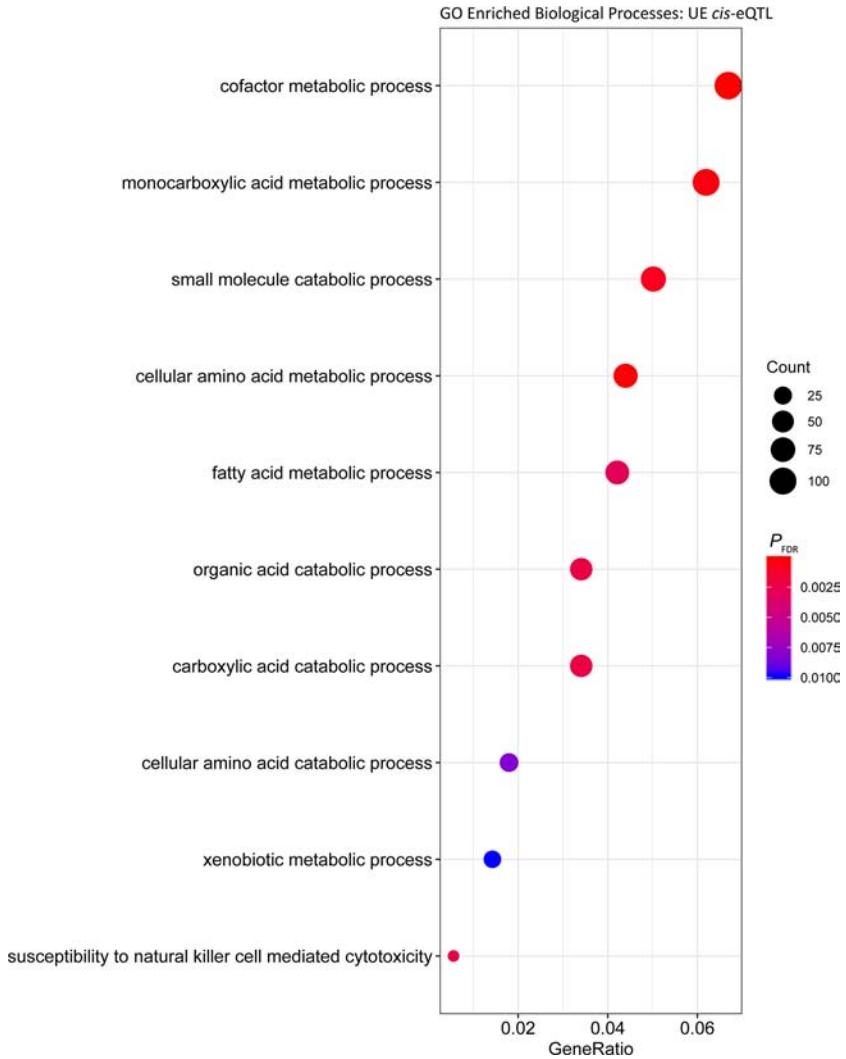


FIGURE 3 | Gene ontology enrichment of biological processes for genes under *cis* regulation in the post-exercise cohort.

Genetic Regulation of Training Relevant Genes

Using 3,405 genes that were differentially expressed in response to training in a sample of 39 Thoroughbreds (Bryan et al, 2017), we examined our results based on eQTL associated with genes within this transcriptional response. Within the UR cohort, 609 of the 3,582 *cis*-eQTL were associated with training response genes. The strongest association was between BIEC2-1061469 and the spindle and expression of the kinetochore associated complex subunit 1 gene (*SKA1*; $P_{FDR} = 9.80 \times 10^{-18}$) (**Table 7**). Of the 1,703 UR *trans*-eQTL, 145 were associated with training response genes. The most significant association was between BIEC2-658237 and *TAL2* ($P_{FDR} = 3.03 \times 10^{-10}$) (**Table 8**).

Within the UE cohort 766 of the 4,992 *cis*-eQTL were associated with training response genes. The most significant *cis*-eQTL association was between UKUL3712 and the interleukin 33 gene (*IL33*; $P_{FDR} = 8.07 \times 10^{-16}$) (**Table 7**). Of the 1,219 UE *trans*-eQTL 90 were associated with genes relevant to training. As with the exercise relevant genes, the strongest UE *trans*-eQTL was between BIEC2-1053404 and *PRDX2* ($P_{FDR} = 1.51 \times 10^{-8}$) (**Table 8**).

DISCUSSION

Using a systems genetics approach we have integrated RNA-seq and genome-wide SNP data for a large cohort of Thoroughbred horses in active race training that were maintained in a single environment. This strategy has allowed us to detect significant *cis* and *trans* eQTL in equine skeletal muscle that are likely to be relevant to an exercise phenotype, adaptation to training, an important and valuable trait in the racing Thoroughbred. A total of 4,992 *cis*-eQTL associated with the expression of 1,922 distinct genes were identified in the UR cohort; and 4,886 *cis*-eQTL associated with the expression of 1,875 genes were identified in

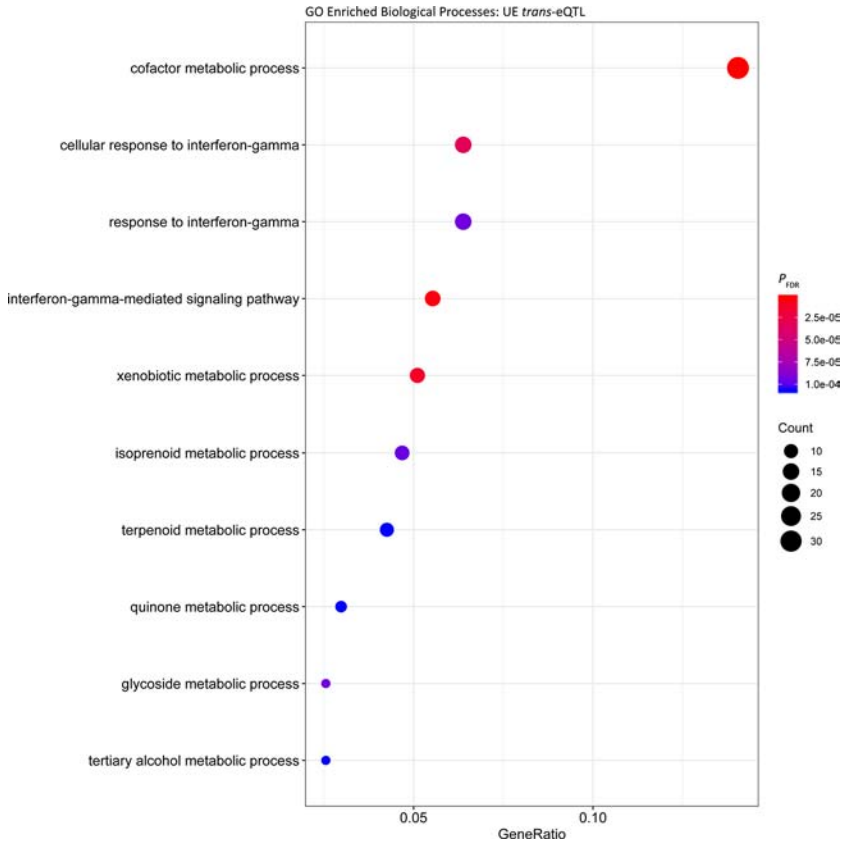


FIGURE 4 | Gene ontology enrichment of biological processes for genes under *trans* regulation in the post-exercise cohort.

TABLE 5 | top ten *cis*-eQTL identified in genes differentially expressed in response to exercise.

SNP	Ensembl Gene ID	Gene	Log_2FC: Exercise[1]	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting											
BIEC2-285235	ENSECAG00000011967	*CNOT11*	0.34	15	8238549	3.00E-15	15	8266315	8276804	-1	-27766
BIEC2-240006	ENSECAG00000017792	*PARN*	0.37	13	30544927	4.42E-15	13	30444515	30591937	1	100412
BIEC2-209001	ENSECAG00000019190	*IFT22*	-0.36	13	9212034	5.33E-15	13	9228359	9233728	-1	-16325
BIEC2-1035081	ENSECAG00000008723	NA	0.47	7	14212942	5.82E-14	7	14574017	14585790	-1	-361075
BIEC2-830969	ENSECAG00000012898	*PNLDC1*	0.55	31	1778298	3.95E-13	31	2140239	2157442	1	-361941
BIEC2-872079	ENSECAG00000024368	*IPO9*	-0.62	30	28911273	1.23E-12	30	28818157	28849624	1	93116
BIEC2-988215	ENSECAG00000018541	*C2CD2L*	-0.59	7	26800468	4.48E-12	7	26815981	26823699	1	-15513
BIEC2-911228	ENSECAG00000024971	*GSTM4*	-0.45	5	57817054	4.38E-11	5	58455048	58459646	-1	-637994
BIEC2-609981	ENSECAG00000011021	*NDUFAF5*	0.41	22	9904659	7.88E-11	22	10479889	10512792	-1	-575230
TBIEC2-57788	ENSECAG00000013327	*SLC24A1*	0.82	1	1.27E+08	9.47E-11	1	126764822	1.27E+08	-1	-3157
Untrained post-exercise											
BIEC2-240006	ENSECAG00000017792	*PARN*	0.37	13	30544927	2.41E-21	13	30444515	30591937	1	100412
BIEC2-209001	ENSECAG00000019190	*IFT22*	-0.36	13	9212034	5.64E-13	13	9228359	9233728	-1	-16325
BIEC2-1082617	ENSECAG00000008479	*SDR16C5*	-0.43	9	27686807	3.39E-12	9	27293107	27307849	1	393700
BIEC2-1035081	ENSECAG00000008723	NA	0.47	7	14212942	6.06E-12	7	14574017	14585790	-1	-361075
TBIEC2-513760	ENSECAG00000000973	NA	-1.82	2	73037320	1.15E-11	2	72462318	72935365	1	575002
BIEC2-770455	ENSECAG00000011697	*TMCC3*	0.57	28	19829349	2.82E-11	28	19803131	19811873	-1	26218
BIEC2-14401	ENSECAG00000010786	*PGAM1*	0.48	1	32009184	6.25E-11	1	31823570	31825848	-1	185614
BIEC2-412279	ENSECAG00000023566	*DHRS9*	1.18	18	48009567	1.04E-10	18	48741376	48752360	1	-731809
BIEC2-1029754	ENSECAG00000019879	*ISCU*	0.50	8	11582493	2.78E-10	8	12361828	12395086	-1	-779335
BIEC2-909217	ENSECAG00000009353	*AGMO*	-0.47	4	48124305	3.15E-10	4	47940706	48126316	-1	183599

[1]*Log_2 fold-change (log_2FC) of expression in response to exercise in Bryan et al., 2017.*

TABLE 6 | top ten *trans*-eQTL identified in genes differentially expressed in response to exercise.

SNP	Ensembl Gene ID	Gene	Log_2FC: Exercise[1]	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting											
BIEC2-658237	ENSECAG00000004995	*TAL2*	-1.67	25	9914109	3.03E-10	25	11725340	11725720	1	-1811231
BIEC2-53419	ENSECAG00000013327	*SLC24A1*	0.82	1	124846027	3.72E-08	1	126764822	126790389	-1	-1918795
BIEC2-58303	ENSECAG00000013327	*SLC24A1*	0.82	1	128180460	2.14E-06	1	126764822	126790389	-1	1415638
BIEC2-913863	ENSECAG00000024971	*GSTM4*	-0.45	5	63531114	5.08E-06	5	58455048	58459646	-1	5076066
BIEC2-959879	ENSECAG00000024971	*GSTM4*	-0.45	5	61035410	3.24E-05	5	58455048	58459646	-1	2580362
BIEC2-496328	ENSECAG00000014441	*ZNF330*	0.56	2	89041634	3.72E-05	2	90360628	90375365	-1	-1318994
BIEC2-311700	ENSECAG00000017760	*ELK3*	0.44	15	53973327	5.26E-05	28	21110553	21148433	1	NA
BIEC2-907346	ENSECAG00000018012	*SELENBP1*	-0.56	5	44309385	6.82E-05	5	45898879	45906904	1	-1589494
BIEC2-59128	ENSECAG00000013327	*SLC24A1*	0.82	1	129427658	1.09E-04	1	126764822	126790389	-1	2662836
BIEC2-60026	ENSECAG00000013327	*SLC24A1*	0.82	1	130739790	1.59E-04	1	126764822	126790389	-1	3974968
Untrained post-exercise											
BIEC2-1053404	ENSECAG00000011541	*PRDX2*	0.53	7	51698577	1.51E-08	7	45573319	45575352	1	6125258
BIEC2-996678	ENSECAG00000011541	*PRDX2*	0.53	7	41340513	2.58E-08	7	45573319	45575352	1	-4232806
BIEC2-658237	ENSECAG00000004995	*TAL2*	-1.67	25	9914109	9.79E-08	25	11725340	11725720	1	-1811231
BIEC2-1028898	ENSECAG00000019879	*ISCU*	0.50	8	10602506	1.29E-06	8	12361828	12395086	-1	-1759322
BIEC2-1028319	ENSECAG00000019879	*ISCU*	0.50	8	10149470	9.98E-06	8	12361828	12395086	-1	-2212358
BIEC2-1051774	ENSECAG00000023096	*TPGS2*	0.41	8	51698367	2.02E-05	8	56940922	56994571	-1	-5242555
BIEC2-420441	ENSECAG00000009547	*WDR12*	0.57	18	76611180	4.93E-05	18	77655352	77674750	-1	-1044172
TBIEC2-1144261	ENSECAG00000008479	*SDR16C5*	-0.43	9	30312051	4.94E-05	9	27293107	27307849	1	3018944
TBIEC2-1149509	ENSECAG00000008479	*SDR16C5*	-0.43	9	40285361	6.34E-05	9	27293107	27307849	1	12992254
BIEC2-1052388	ENSECAG00000023096	*TPGS2*	0.41	8	53636613	1.07E-04	8	56940922	56994571	-1	-3304309

[1]*Log_2 fold-change (log_2FC) of expression in response to exercise in Bryan et al., 2017.*

TABLE 7 | top *cis*-eQTL identified in genes differentially expressed in response to training.

SNP	Ensembl Gene ID	Gene	Log_2FC: Training[1]	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting											
BIEC2-1061469	ENSECAG00000023484	*SKA1*	0.41	8	69117228	9.80E-18	8	68814638	68828326	1	302590
BIEC2-1061072	ENSECAG00000023484	*SKA1*	0.41	8	68498709	3.57E-17	8	68814638	68828326	1	-315929
TBIEC2-1028240	ENSECAG00000022151	*MATK*	0.58	7	2020098	5.03E-13	7	2272120	2277087	-1	-252022
BIEC2-974177	ENSECAG00000022151	*MATK*	0.58	7	1546395	1.12E-12	7	2272120	2277087	-1	-725725
BIEC2-872079	ENSECAG00000024368	*IPO9*	-0.66	30	28911273	1.23E-12	30	28818157	28849624	1	93116
BIEC2-988215	ENSECAG00000018541	*C2CD2L*	-0.43	7	26800468	4.48E-12	7	26815981	26823699	1	-15513
BIEC2-472999	ENSECAG00000021368	*FGGY*	0.37	2	732213	7.94E-12	2	549904	834638	-1	182309
BIEC2-472598	ENSECAG00000015014	*ACADL*	0.44	2	248866	1.08E-11	2	124580	165039	-1	124286
BIEC2-609981	ENSECAG00000011021	*NDUFAF5*	0.43	22	9904659	7.88E-11	22	10479889	10512792	-1	-575230
TBIEC2-57788	ENSECAG00000013327	*SLC24A1*	0.74	1	1.27E+08	9.47E-11	1	1.27E+08	1.27E+08	-1	-3157
Untrained post-exercise											
UKUL3712	ENSECAG00000010475	*IL33*	0.74	23	27442828	8.07E-16	23	27442768	27454175	1	60
BIEC2-974177	ENSECAG00000022151	*MATK*	0.58	7	1546395	1.96E-14	7	2272120	2277087	-1	-725725
TBIEC2-1028240	ENSECAG00000022151	*MATK*	0.58	7	2020098	7.95E-14	7	2272120	2277087	-1	-252022
BIEC2-1082617	ENSECAG00000008479	*SDR16C5*	-0.33	9	27686807	3.39E-12	9	27293107	27307849	1	393700
BIEC2-472598	ENSECAG00000015014	*ACADL*	0.44	2	248866	8.29E-12	2	124580	165039	-1	124286
TBIEC2-513760	ENSECAG00000000973	NA	-0.85	2	73037320	1.15E-11	2	72462318	72935365	1	575002
BIEC2-360074	ENSECAG00000014129	*TMEM42*	0.34	16	42029863	2.23E-10	16	41642464	41643652	-1	387399
BIEC2-472999	ENSECAG00000021368	*FGGY*	0.37	2	732213	2.74E-10	2	549904	834638	-1	182309
BIEC2-1029754	ENSECAG00000019879	*ISCU*	0.44	8	11582493	2.78E-10	8	12361828	12395086	-1	-779335
BIEC2-136659	ENSECAG00000023268	*CALHM5*	0.54	10	64788579	3.77E-10	10	65017017	65021618	1	-228438

[1]*Log_2 fold-change (log_2FC) of expression in response to training in Bryan et al., 2017.*

the UE cohort. Fewer *trans*-eQTL were detected (UR: 1,703; UE: 1,219), which is consistent with previous studies, and likely due to the greater statistical power required to identify *trans*-eQTL (Westra and Franke, 2014).

The gene with the most significant association with a *cis*-eQTL in the UR and UE cohorts was *ENTR1* (**Table 3**, UR: $P_{FDR} = 3.81 \times 10^{-27}$, UE: $P_{FDR} = 1.66 \times 10^{-24}$). The ENTR1 protein is involved in cellular transport of cargo proteins from the endosome to the

TABLE 8 | top *trans*-eQTL identified in genes differentially expressed in response to training.

SNP	Ensembl Gene ID	Gene	Log_2FC: Training[1]	SNP Chr	SNP Pos	P_{FDR}	Gene Chr	TSS	Gene End	Strand	Distance from TSS
Untrained resting											
BIEC2-658237	ENSECAG00000004995	*TAL2*	-1.63	25	9914109	3.03E-10	25	11725340	11725720	1	-1811231
BIEC2-1118794	ENSECAG00000023484	*SKA1*	0.41	8	67640813	4.93E-10	8	68814638	68828326	1	-1173825
BIEC2-53419	ENSECAG00000013327	*SLC24A1*	0.74	1	124846027	3.72E-08	1	126764822	126790389	-1	-1918795
BIEC2-58303	ENSECAG00000013327	*SLC24A1*	0.74	1	128180460	2.14E-06	1	126764822	126790389	-1	1415638
BIEC2-59128	ENSECAG00000013327	*SLC24A1*	0.74	1	129427658	1.09E-04	1	126764822	126790389	-1	2662836
BIEC2-196080	ENSECAG00000017776	*PITPNM1*	0.38	12	25661692	1.28E-04	12	27228025	27239903	-1	-1566333
BIEC2-542299	ENSECAG00000021368	*FGGY*	0.37	20	7321704	1.34E-04	2	549904	834638	-1	NA
BIEC2-60026	ENSECAG00000013327	*SLC24A1*	0.74	1	130739790	1.59E-04	1	126764822	126790389	-1	3974968
BIEC2-580450	ENSECAG00000011021	*NDUFAF5*	0.43	22	9444819	2.10E-04	22	10479889	10512792	-1	-1035070
BIEC2-328903	ENSECAG00000020701	*CAV3*	-0.37	16	6817902	2.43E-04	16	8001463	8011813	-1	-1183561
Untrained post-exercise											
BIEC2-1053404	ENSECAG00000011541	*PRDX2*	0.59	7	51698577	1.51E-08	7	45573319	45575352	1	6125258
BIEC2-996678	ENSECAG00000011541	*PRDX2*	0.59	7	41340513	2.58E-08	7	45573319	45575352	1	-4232806
BIEC2-658237	ENSECAG00000004995	*TAL2*	-1.63	25	9914109	9.79E-08	25	11725340	11725720	1	-1811231
BIEC2-1028898	ENSECAG00000019879	*ISCU*	0.45	8	10602506	1.29E-06	8	12361828	12395086	-1	-1759322
BIEC2-344350	ENSECAG00000014129	*TMEM42*	0.34	16	43032493	4.09E-06	16	41642464	41643652	-1	1390029
BIEC2-1028319	ENSECAG00000019879	*ISCU*	0.45	8	10149470	9.98E-06	8	12361828	12395086	-1	-2212358
BIEC2-130408	ENSECAG00000023268	*CALHM5*	0.54	10	67512852	2.28E-05	10	65017017	65021618	1	2495835
TBIEC2-1144261	ENSECAG00000008479	*SDR16C5*	-0.33	9	30312051	4.94E-05	9	27293107	27307849	1	3018944
BIEC2-809036	ENSECAG00000005431	*CLRN2*	0.67	3	107506494	6.34E-05	3	106028106	106037626	-1	1478388
TBIEC2-1149509	ENSECAG00000008479	*SDR16C5*	-0.33	9	40285361	6.34E-05	9	27293107	27307849	1	12992254

[1]*Log_2 fold-change (log_2FC) of expression in response to training in Bryan et al., 2017.*

Golgi apparatus or for degradation in the lysosome (McGough et al., 2014) and has been suggested to play a role in cytokinesis (Hagemann et al., 2013). In a study where ENTR1 protein expression was blocked by RNA interference, there was a decrease in solute carrier family 2 member 1 glucose transporter protein (SLC2A1; previously known as GLUT1) (McGough et al., 2014). When examining whether this was due to increased SLC2A1 degradation, there was no evidence of increased transport of SLC2A1 to the lysosome. It was hypothesised that the decrease in SLC2A1 was mediated through regulation of transcription by ENTR1. SLC2A1 is responsible for approximately 30–40% of the glucose uptake in skeletal muscle, with the remainder transported through GLUT4 (Zisman et al., 2000; Rudich et al., 2003). As opposed to GLUT4 which is primarily expressed in skeletal muscle, SLC2A1 is widely expressed and is highly expressed on erythrocyte membranes (Krook et al., 2004). The control of SLC2A1 by ENTR1 in the context of the equine athlete is intriguing to speculate since SLC2A1 is expressed within equine lamellar tissue, and its expression is increased in hyperinsulinemia, therefore may play a role in the pathophysiology of equine laminitis (Campolo et al., 2016). *SLC2A1* is also differentially expressed in response to hypoxia, this has also been shown in equine chondrocytes *in vitro* after exposure to cobalt chloride (to mimic hypoxia) and in chondrocytes from osteoarthritis cases (Peansukmanee et al., 2009).

The most significant *trans* association in the UR cohort was between BIEC2-526896 and expression of the DEAH-box helicase 16 gene (*DHX16*) (**Table 4**). *DHX16* is an RNA helicase and is involved in regulation of translation and pre-mRNA splicing (Gencheva et al., 2010; Putiri and Pelegri, 2011). The gene located closest to BIEC2-526896 is the olfactory receptor family 12 subfamily D member 3 gene (*OR12D3*) with the TSS located 96.5 kb from the SNP. However, the zinc finger protein 311 gene (ZNF311) also relatively close to BIEC2-526896 (792.5 kb)(Consortium, 2017). *ZNF311* has previously been associated with telomere length in heterozygous ataxia-talengiectasia mutated (*ATM*) gene patients (Renault et al., 2017). As a member of the a krueppel c2h2-type zinc-finger protein family it is likely a transcription factor and has been associated with Biological Processes such as 'regulation of transcription, DNA templated' and 'regulation of transcription by RNA polymerase II'(Consortium, 2017). The *trans* association between BIEC2-526896 and *DHX16* expression may therefore be mediated *via* the gene regulatory function of *ZNF311*.

The most significant *trans*-eQTL in the UE cohort was UKUL2765 and expression of the methylcrotonoyl-CoA carboxylase 2 gene (*MCCC2*) (**Table 4**). *MCCC2* encodes a subunit of 3-methylcrotonyl-CoA carboxylase (MCC), an enzyme which catabolises leucine (Stadler et al., 2005). Mutations within *MCCC2* have been found to result in MCC deficiency, which has varying implications for patients from no symptoms at all to death in early infancy (Fonseca et al., 2016). To date studies have yet to discern mutations which result in more or less severe disease phenotypes (Gallardo et al., 2001; Stadler et al., 2006). In terms of muscle physiology, *MCCC2* has been shown to be highly expressed in skeletal muscle of the red seabream fish (*Pagrus major*), which is likely due to high levels of protein metabolism within skeletal muscle (Abe et al., 2004). The TSS of the jumonji domain containing 4 gene (*JMJD4*) is located 71 bp from UKUL2765. The JMJD4 protein catalyses the hydroxylation of translation termination factor eRF1 lysine 63, which in turn enables the correct termination of translation and maintenance of translational fidelity (Feng et al., 2014). It is possible that the variation proximal to *JMJD4* is influencing expression of *JMJD4*,

in turn altering expression of *MCCC2*. However, from the data available only one significant *cis*-eQTL for *JMJD4* was detected in the UR cohort and this was BIEC2-277622 located 257.8 kb downstream of the TSS ($P_{FDR} = 6.58 \times 10^{-5}$). Therefore it is not clear if UKUL2765 is tagging variation influencing *JMJD4* expression and mediating its influence on *MCCC2* through *JMJD4*.

Examination of eGenes previously shown to be involved in the skeletal muscle transcriptional response to exercise and training demonstrated that *TAL2* exhibited the most significant *trans*-eQTL in the UR cohort (BIEC2-658237; **Table 6**) and that this *trans*-eQTL was also highly significant in the UE cohort ($P_{FDR} = 9.80 \times 10^{-8}$; **Table 4**). *TAL2* encodes a basic-helix-loop-helix transcription factor (Xia et al., 1991; Langlands et al., 1997). Deletion of *TAL2* in mice has been shown to cause severe disruption of the development of the central nervous system, with new-born mice dying shortly post-partum (Bucher et al., 2000). *TAL2* has been shown to be vital for the development of gamma-aminobutyric acid (GABA, inhibitory neurotransmitter) signalling neurons in the developing midbrain, showing highly regulated and coordinated expression (Achim et al., 2013). When expression of *TAL2* was inhibited, neurons more closely resembled an excitatory glutamatergic phenotype (Achim et al., 2013). In terms of application in racing performance, GABA has previously been used as a calming agent in Thoroughbred racehorses, although it was banned from use in 2012. The GABA type A receptor associated protein like 1 gene (*GABARAPL1*) was also identified as a key regulator in the skeletal muscle transcriptional response to exercise (Bryan et al., 2017). In addition, we have previously reported functional enrichment of pathways related to neurodegenerative disorders in the transcriptional response to exercise (Bryan et al., 2017). Given the role of *TAL2* in GABAergic neuronal fate, this suggests a potential role for *TAL2* in the coordination of the response to exercise. These results suggest that the role of genes associated with neuronal differentiation and disease in the context of muscle and exercise warrants further investigation.

To identify common biological functions within genes identified under *cis* or *trans* regulation, enrichment analysis of Biological Processes among the gene sets was performed. Among the *cis* eGenes detected in both the UR and UE cohort, as well as *trans* eGenes in the UR cohort there was significant enrichment of cofactor metabolic processes (GO:0051186, **Tables S1**, **S3**, and **S4**). Cofactor metabolic process is defined as chemical reactions and pathways requiring the activity of an inorganic cofactor, such as an ion, or an organic coenzyme for the activity of an enzyme or other functional protein. Genes within this cluster were related to metabolism and substrate utilisation, including vitamin and mineral binding and synthesis such as: selenium (selenium binding protein 1 gene, *SELENBP1*; and selenoprotein T gene, *SELENOT*), molybdenum (molybdenum cofactor sulfurase gene, *MOCOS*) and thiamine (thiamine triphosphatase gene, *THTPA*). Consequently, variation in the expression of genes associated with nutrient binding may lead to variation in the ability of horses to utilise such nutrients. In this regard, abundance of selenoprotein gene transcripts has been used to identify dietary requirements for selenium in rats and turkeys (Barnes et al., 2009; Taylor and Sunde, 2017). Given the inter-animal variation in expression observed for genes relevant to substrate binding, it may be possible to use this information to evaluate nutrient requirements for individual horses, or whether expression of these genes can be modulated through diet.

Many of these genes have also been shown to have functions relevant to exercise, and variation within the expression of these genes may underpin variation in athletic performance. For example, the selenium binding protein 1 gene (*SELENBP1*) is significantly downregulated in response to exercise ($\log_2$FC = −0.56; $P_{FDR} = 3.71 \times 10^{-11}$) (Bryan et al., 2017). In both normal and cancerous human cells *SELENBP1* has been shown to be highly variable in expression (Yang and Sytkowski, 1998). Functionally, the SELENBP1 protein has been shown to be involved in many cellular processes including detoxification (Ishii et al., 1996), cytoskeletal outgrowth (Miyaguchi, 2004) and regulation of reduction and oxidation within the cell (Jamba et al., 1997). *SELENBP1* was found be differentially expressed in blood in response to administration of human recombinant erythropoietin in human endurance athletes (Durussel et al., 2016; Wang et al., 2017), suggesting a potential role in haematopoiesis and its regulation. In the UE cohort; the DDB1 and CUL4 associated factor 12 (*DCAF12*) and guanosine monophosphate reductase (*GMPR*) genes both exhibited significant *cis*-eQTL (*DCAF12*: $P_{FDR} = 0.02$; *GMPR*: $P_{FDR} = 4.17 \times 10^{-3}$). These genes, in addition to *SELENBP1*, were also shown by Wang et al. (2017) to be differentially expressed in blood in response to human recombinant erythropoietin. Variation in the expression of these genes may therefore potentially underpin variation in haematological phenotypes in horses, which may in turn influence traits relevant to aerobic capacity. It is also noteworthy that selenium deficiency has been associated with significant myopathy (White muscle disease) (Lofstedt, 1997; Delesalle et al., 2017) and reduced exercise tolerance in horses (Brady et al., 1978; Avellini et al., 1999). In addition, selenoproteins have been shown to be involved in several metabolic pathways and the response to oxidative stress in muscle (Rederstorff et al., 2006). These findings suggest an important role for selenium, and its associated biochemical machinery, in the correct functioning of skeletal muscle and muscle metabolism. This highlights the importance of selenium in the context of exercise and provides a potential role for variation in expression of genes relevant to selenium metabolism in determining metabolic function within the muscle.

The cofactor metabolic process cluster also contained genes relevant to mitochondrial function and oxidative phosphorylation. These included genes within the coenzyme Q synthesis pathway: coenzyme Q3 hydroxylase (*COQ3*), coenzyme Q7 (*COQ7*) and coenzyme Q8A (*COQ8A*). The coenzyme Q complex is a critical component of the electron transport chain during oxidative phosphorylation, moving electrons from complexes I and II to complex III (Lenaz, 1985; Turunen et al., 2004; Stefely and Pagliarini, 2017). COQ7 and COQ8A are required for coenzyme Q biosynthesis (Mollet et al., 2008; Stefely et al., 2016). Human patients with *COQ8A* mutations suffered seizures and other neurological symptoms and showed reduced coenzyme Q within skeletal muscle (Jacobsen et al.; Mollet et al., 2008). An eQTL for *COQ8A* in skeletal muscle has already been identified in horses, with a 227 bp SINE insertion in the promotor region of *MSTN* (g.66495326_66495327ins227) on ECA18 associated with

increased expression of *COQ8A* (previously known as *ADCK3*) in Thoroughbreds (Rooney et al., 2017). However it should be noted that this increase in COQ8A expression did not appear to accompany an COQ8A protein abundance, with no difference in COQ8A protein abundance across genotypes (Rooney et al., 2017). This may be due to COQ8A having a regulatory role in coenzyme Q biosynthesis (Acosta et al., 2016). Electron transport chain complex activity assays, as well as assays using the exogenous application of ubiquinone, suggested a difference in the abundance of coenzyme Q across genotypes at this locus (Rooney et al., 2017). Suggesting variation at this SINE insertion is associated with *COQ8A* expression as well as coenzyme Q abundance. Therefore eQTL in the current study associated with *COQ8A* (**Figure S4**), and indeed other genes within the coenzyme Q biosynthetic pathway, may result in variation in synthesis of the coenzyme Q complex and have downstream implications for mitochondrial function.

We have for the first time systematically catalogued eQTL in equine skeletal muscle, both at rest and post-exercise. Previous investigations of eQTL in skeletal muscle have focussed primarily on human T2D (Sharma et al., 2011; Keildson et al., 2014a; Sajuthi et al., 2016; Langefeld et al., 2018) and meat quality traits in production animals (Ponsuksili et al., 2015; Gonzalez-Prendes et al., 2017; Pampouille et al., 2018; Gonzalez-Prendes et al., 2019; Velez-Irizarry et al., 2019). Our investigation of eQTL in the context of skeletal muscle and exercise present some of the only work to-date in this area (Kelly et al., 2014). Our work utilised linear models to detect associations between SNPs and gene expression as quantified by RNA-seq. It should be noted at this point that there are potential biases introduced in terms of the high number of related individuals within the cohort, future work could utilise more sophisticated techniques such as allele specific expression, where transcripts are mapped back to the maternal and paternal chromosomes and the expression of the maternal and paternal transcripts can be compared (Chamberlain et al., 2015). This would be particularly useful in our cohort given the high number of offspring by a small number of sires. Within the cohort there is also some variation in the amount of training prior to sampling, this was kept to a minimum by our sampling criteria. However, extending the study to incorporate trained horses and utilise variation in prior training by modelling the transcriptional training response could provide information on regulation of the training responsive transcriptome.

Our current results provide novel information concerning the regulation of gene expression in horses and can provide a framework for interpreting future GWAS of athletic and performance traits in Thoroughbreds. In terms of future applications of these results, the identification of quantitative trait transcripts (QTT) for athletic traits characterised in our cohort could be used to detect associations between a SNP, variation in expression of a QTT and a trait of interest. Thus giving a fuller picture of genetic variation contributing to traits of interest. An example may be detecting loci and QTT involved in the response to exercise and training, which has previously been shown to be highly heritable in humans (Timmons et al., 2010; Bouchard et al., 2011). The use of systems genetics approaches that integrate differential gene expression with genome variation represent an excellent strategy for dissecting the genetic architecture of complex anatomical and physiological traits.

DATA AVAILABILITY STATEMENT

The RNA-seq datasets analyzed for this study can be found in EBI ArrayExpress https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5447/. SNP datasets will not be made publicly available as the data were generated from privately owned horses, with a legal commitment to confidentiality. Researchers may request access to the data and consideration will be given to individuals following the conclusion of a confidentiality agreement. Requests should be made to the UCD Technology Transfer Office (https://www.ucd.ie/innovation/knowledge-transfer/).

ETHICS STATEMENT

University College Dublin Animal Research Ethics Committee approval (AREC-P-12-55-Hill), a licence from the Department of Health (B100/3525) and informed owner consent were obtained.

AUTHOR CONTRIBUTIONS

GF performed computations and functional analyses. KB and PM assisted in analysis and pipeline development. CM, KG and GF performed biopsy sample collections. JB and GF prepared RNA. GF wrote the manuscript in close consultation with EH, LK and DM. All authors were involved in study design, implementation of the research and preparation of the manuscript.

FUNDING

This research was funded by Science Foundation Ireland (SFI/11/PI/1166 and 18/TIDA/6019).

ACKNOWLEDGMENTS

We would like to thank J.S. Bolger for access to his horses, and staff at Glebe House stables for their assistance, particularly B. O'Connor and P. O'Donovan. This research was conducted with the financial support of Science Foundation Ireland (grant no. SFI/11/PI/1166 and 18/TIDA/6019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01215/full#supplementary-material

FIGURE S1 | Principal component analysis of quantile-normalised log counts of RNA-seq transcripts and coloured by sample type: untrained resting or untrained post-exercise.

FIGURE S2 | Boxplot of ENTR1 expression (log2 quantile-normalised counts) across BIEC2-707785 genotypes in untrained resting samples.

FIGURE S3 | Boxplot of ENTR1 expression (log2 quantile-normalised counts) across BIEC2-707785 genotypes in untrained post-exercise samples.

FIGURE S4 | Boxplot of COQ8A expression (log2 quantile-normalised counts) across BIEC2-417075 genotypes in untrained post-exercise samples.

REFERENCES

Abe, S., Chamnan, C., Miyamoto, K., Minamino, Y., and Nouda, M. (2004).Isolation and identification of 3-methylcrotonyl coenzyme A carboxylase cDNAs and pyruvate carboxylase, and their expression in red seabream (Pagrus major) organs. *Mar. Biotechnol. (N. Y.)* 6 (6), 527–540. doi: 10.1007/s10126-004-4203-x

Achim, K., Peltopuro, P., Lahti, L., Tsai, H. H., Zachariah, A., Astrand, M., et al. (2013). The role of Tal2 and Tal1 in the differentiation of midbrain GABAergic neuron precursors. *Biol. Open* 2 (10), 990–997. doi: 10.1242/bio.20135041

Acosta, M. J., Vazquez Fonseca, L., Desbats, M. A., Cerqua, C., Zordan, R., Trevisson, E., et al. (2016). Coenzyme Q biosynthesis in health and disease. *Biochim. Biophys. Acta (BBA) - Bioenerg.* 1857 (8), 1079–1085. doi: 10.1016/j.bbabio.2016.03.036

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Ref. Source*.

Arany, Z. (2008). PGC-1 coactivators and skeletal muscle adaptations in health and disease. *Curr. Opin. Genet. Dev.* 18 (5), 426–434. doi: 10.1016/j.gde.2008.07.018

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Avellini, L., Chiaradia, E., and Gaiti, A. (1999). Effect of exercise training, selenium and vitamin E on some free radical scavengers in horses (Equus caballus). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 123 (2), 147–154. doi: 10.1016/s0305-0491(99)00045-0

Barnes, K. M., Evenson, J. K., Raines, A. M., and Sunde, R. A. (2009). Transcript analysis of the selenoproteome indicates that dietary selenium requirements of rats based on selenium-regulated selenoprotein mRNA levels are uniformly less than those based on glutathione peroxidase activity. *J. Nutr.* 139 (2), 199–206. doi: 10.3945/jn.108.098624

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57 (1), 289–300. doi: 10.2307/2346101

Bolstad, B. (2017). PreprocessCore: a collection of pre-processing functions in *R package version*, (https://github.com/bmbolstad/preprocessCore).

Bouchard, C., Sarzynski, M. A., Rice, T. K., Kraus, W. E., Church, T. S., Sung, Y. J., et al. (2011). Genomic predictors of the maximal O-2 uptake response to standardized exercise training programs. *J. Appl. Physiol.* 110 (5), 1160–1170. doi: 10.1152/japplphysiol.00973.2010

Brady, P. S., Ku, P. K., and Ullrey, D. E. (1978). Lack of effect of selenium supplementation on the response of the equine erythrocyte glutathione system and plasma enzymes to exercise. *J. Anim. Sci.* 47 (2), 492–496. doi: 10.2527/jas1978.472492x

Brown, C. D., Mangravite, L. M., and Engelhardt, B. E. (2013). Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PloS Genet.* 9 (8), e1003649. doi: 10.1371/journal.pgen.1003649

Bryan, K., McGivney, B. A., Farries, G., McGettigan, P. A., McGivney, C. L., Gough, K. F., et al. (2017). Equine skeletal muscle adaptations to exercise and training: evidence of differential regulation of autophagosomal and mitochondrial components. *BMC Genomics* 18 (1), 595. doi: 10.1186/s12864-017-4007-9

Bucher, K., Sofroniew, M. V., Pannell, R., Impey, H., Smith, A. J., Torres, E. M., et al. (2000). The T cell oncogene Tal2 is necessary for normal development of the mouse brain. *Dev. Biol.* 227 (2), 533–544. doi: 10.1006/dbio.2000.9920

Campolo, A., de Laat, M. A., Keith, L., Gruntmeir, K. J., and Lacombe, V. A. (2016). Prolonged hyperinsulinemia affects metabolic signal transduction markers in a tissue specific manner. *Domest. Anim. Endocrinol.* 55, 41–45. doi: 10.1016/j.domaniend.2015.11.001

Carlson M (2019). *org.Hs.eg.db: Genome wide annotation for Human*. R package version 3.8.2.

Chamberlain, A. J., Vander Jagt, C. J., Hayes, B. J., Khansefid, M., Marett, L. C., Millen, C. A., et al. (2015). Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* 16, 993. doi: 10.1186/s12864-015-2174-0

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi: 10.1186/s13742-015-0047-8

Consortium, U. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–d169. doi: 10.1093/nar/gkw1099

Corbin, L. J., Kranis, A., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., et al. (2014). The utility of low-density genotyping for imputation in the Thoroughbred horse. *Genet. Sel. Evol.* 46, 9. doi: 10.1186/1297-9686-46-9

Delesalle, C., de Bruijn, M., Wilmink, S., Vandendriessche, H., Mol, G., Boshuizen, B., et al. (2017). White muscle disease in foals: focus on selenium soil content. A case series. *BMC Vet. Res.* 13 (1), 121–121. doi: 10.1186/s12917-017-1040-5

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4 (8), 1184–1191. doi: 10.1038/nprot.2009.97

Durussel, J., Haile, D. W., Mooses, K., Daskalaki, E., Beattie, W., Mooses, M., et al. (2016). Blood transcriptional signature of recombinant human erythropoietin administration and implications for antidoping strategies. *Physiol. Genomics* 48 (3), 202–209. doi: 10.1152/physiolgenomics.00108.2015

Eivers, S. S., McGivney, B. A., Fonseca, R. G., MacHugh, D. E., Menson, K., Park, S. D., et al. (2010). Alterations in oxidative gene expression in equine skeletal muscle following exercise and training. *Physiol. Genomics* 40 (2), 83–93. doi: 10.1152/physiolgenomics.00041.2009

Eivers, S. S., McGivney, B. A., Gu, J., MacHugh, D. E., Katz, L. M., and Hill, E. W. (2012). PGC-1 alpha encoded by the PPARGC1A gene regulates oxidative energy metabolism in equine skeletal muscle during exercise. *Anim. Genet.* 43 (2), 153–162. doi: 10.1111/j.1365-2052.2011.02238.x

Farries, G., Gough, K. F., Parnell, A. C., McGivney, B. A., McGivney, C. L., McGettigan, P. A., et al. (2019). Analysis of genetic variation contributing to measured speed in Thoroughbreds identifies genomic regions involved in the transcriptional response to exercise. *Anim. Genet.* 50, 670–685. doi: 10.1111/age.12848

Feng, T., Yamamoto, A., Wilkins, S. E., Sokolova, E., Yates, L. A., Munzel, M., et al. (2014). Optimal translational termination requires C4 lysyl hydroxylation of eRF1. *Mol. Cell* 53 (4), 645–654. doi: 10.1016/j.molcel.2013.12.028

Fonseca, H., Azevedo, L., Serrano, C., Sousa, C., Marcao, A., and Vilarinho, L. (2016). 3-Methylcrotonyl-CoA carboxylase deficiency: Mutational spectrum derived from comprehensive newborn screening. *Gene* 594 (2), 203–210. doi: 10.1016/j.gene.2016.09.003

Gallardo, M. E., Desviat, L. R., Rodríguez, J. M., Esparza-Gordillo, J., Pérez-Cerdá, C., Pérez, B., et al. (2001). The molecular basis of 3-methylcrotonylglycinuria, a disorder of leucine catabolism. *Am. J. Hum. Genet.* 68 (2), 334–346. doi: 10.1086/318202

Gencheva, M., Lin, T. Y., Wu, X., Yang, L., Richard, C., Jones, M., et al. (2010). Nuclear retention of unspliced pre-mRNAs by mutant DHX16/hPRP2, a spliceosomal DEAH-box protein. *J. Biol. Chem.* 285 (46), 35624–35632. doi: 10.1074/jbc.M110.122309

Gonzalez-Prendes, R., Quintanilla, R., and Amills, M. (2017). Investigating the genetic regulation of the expression of 63 lipid metabolism genes in the pig skeletal muscle. *Anim. Genet.* 48 (5), 606–610. doi: 10.1111/age.12586

Gonzalez-Prendes, R., Marmol-Sanchez, E., Quintanilla, R., Castello, A., Zidi, A., and Ramayo-Caldas, Y. (2019). About the existence of common determinants of gene expression in the porcine liver and skeletal muscle. *BMC Genomics* 20 (1), 518. doi: 10.1186/s12864-019-5889-5

Gunn, H. M. (1987). Muscle, bone and fat proportions and muscle distribution of thoroughbreds and other horses in *Equine exercise physiology 2: Proceedings of the second international conference on equine exercise physiology*. Eds. Gillespie, J. R., and Robinson, N. E.(San Diego, California), 253–264. August 7-11, 1986.

Hagemann, N., Ackermann, N., Christmann, J., Brier, S., Yu, F., and Erdmann, K. S. (2013). The serologically defined colon cancer antigen-3 interacts with the protein tyrosine phosphatase PTPN13 and is involved in the regulation of cytokinesis. *Oncogene* 32 (39), 4602–4613. doi: 10.1038/onc.2012.485

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* 106 (23), 9362–9367. doi: 10.1073/pnas.0903103106

Ishii, Y., Hatsumura, M., Ishida, T., Ariyoshi, N., and Oguri, K. (1996). Significant induction of a 54-kDa protein in rat liver with homologous alignment to mouse selenium binding protein by a coplanar polychlorinated biphenyl, 3,4,5,3',4'-pentachlorobiphenyl and 3-methylcholanthrene. *Toxicol. Lett.* 87 (1), 1–9. doi: 10.1016/0378-4274(96)03668-5

Jacobsen, J. C., Whitford, W., Swan, B., Taylor, J., Love, D. R., Hill, R., et al. *Compound Heterozygous Inheritance of Mutations in Coenzyme Q8A Results in Autosomal Recessive Cerebellar Ataxia and Coenzyme Q10 Deficiency in a Female Sib-Pair*. Berlin, Heidelberg:Springer Berlin Heidelberg, 1–6. doi: 10.1007/8904_2017_73

Jamba, L., Nehru, B., and Bansal, M. P. (1997). Redox modulation of selenium binding proteins by cadmium exposures in mice. *Mol. Cell Biochem.* 177 (1-2), 169–175. doi: 10.1023/a:1006869623864

Jansen, R. C., and Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends Genet.* 17 (7), 388–391. doi: 10.1016/s0168-9525(01)02310-1

Kasprzyk, A. (2011). BioMart: driving a paradigm change in biological data management. *Database* 2011, bar049. doi: 10.1093/database/bar049

Kayar, S. R., Hoppeler, H., Lindstedt, S. L., Claassen, H., Jones, J. H., Essengustavsson, B., et al. (1989). Total muscle mitochondrial volume in relation to aerobic capacity of horses and steers. *Pflugers Archiv-European J. Physiol.* 413 (4), 343–347. doi: 10.1007/bf00584481

Kearns, C. F., McKeever, K. H., and Abe, T. (2002). Overview of horse body composition and muscle architecture: Implications for performance. *Vet. J.* 164 (3), 224–234. doi: 10.1053/tvjl.2001.0702

Keildson, S., Fadista, J., Ladenvall, C., Hedman, A. K., Elgzyri, T., Small, K. S., et al. (2014a). Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes* 63 (3), 1154–1165. doi: 10.2337/db13-1301

Keildson, S., Fadista, J., Ladenvall, C., Hedman, A. K., Elgzyri, T., Small, K. S., et al. (2014b). Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes* 63 (3), 1154–1165.

Kelly, S. A., Nehrenberg, D. L., Hua, K., Garland, T.Jr., and Pomp, D. (2014). Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. *Physiol. Genomics* 46 (16), 593–601. doi: 10.1152/physiolgenomics.00023.2014

Kraniou, G. N., Cameron-Smith, D., and Hargreaves, M. (2006). Acute exercise and GLUT4 expression in human skeletal muscle: influence of exercise intensity. *J. Appl. Physiol. (1985)* 101 (3), 934–937. doi: 10.1152/japplphysiol.01489.2005

Krook, A., Wallberg-Henriksson, H., and Zierath, J. R. (2004). Sending the signal: molecular mechanisms regulating glucose uptake. *Med. Sci. Sports Exerc.* 36 (7), 1212–1217. doi: 10.1249/01.mss.0000132387.25853.3b

Langefeld, C. D., Comeau, M. E., Sharma, N. K., Bowden, D. W., Freedman, B. I., and Das, S. K. (2018). Transcriptional regulatory mechanisms in adipose and muscle tissue associated with composite glucometabolic phenotypes. *Obesity (Silver Spring)* 26 (3), 559–569. doi: 10.1002/oby.22113

Langlands, K., Yin, X., Anand, G., and Prochownik, E. V. (1997). Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J. Biol. Chem.* 272 (32), 19785–19793. doi: 10.1074/jbc.272.32.19785

Lenaz, G. ed. (1985). Coenzyme Q: Biochemistry, Bioenergetics, and Clinical Applications of Ubiquinone. John Wiley and Sons Incorporated.

Levine, M. (1999). The origins of horse husbandry on the eurasian steppe in *Late prehistoric exploitation of the eurasian steppe*. Eds.Levine, M., Rassamakin, Y., Kislenko, A., and Tatarintseva, N. (Cambridge:McDonald Institute for Archaeological Research), 5–58.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656

Lofstedt, J. (1997).White muscle disease of foals. *Vet. Clin. North Am. Equine Pract.* 13 (1), 169–185. doi: 10.1016/s0749-0739(17)30262-6

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi: 10.1038/ng.2653

Mason, C. C., Hanson, R. L., Ossowski, V., Bian, L., Baier, L. J., Krakoff, J., et al. (2011). Bimodal distribution of RNA expression levels in human skeletal muscle tissue. *BMC Genomics* 12. doi: 10.1186/1471-2164-12-98

McGivney, B. A., Eivers, S. S., MacHugh, D. E., MacLeod, J. N., O'Gorman, G. M., Park, S. D., et al. (2009). Transcriptional adaptations following exercise in thoroughbred horse skeletal muscle highlights molecular mechanisms that lead to muscle hypertrophy. *BMC Genomics* 10, 638. doi: 10.1186/1471-2164-10-638

McGivney, B. A., McGettigan, P. A., Browne, J. A., Evans, A. C., Fonseca, R. G., Loftus, B. J., et al. (2010). Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics* 11, 398. doi: 10.1186/1471-2164-11-398

McGough, I. J., Steinberg, F., Gallon, M., Yatsu, A., Ohbayashi, N., Heesom, K. J., et al. (2014). Identification of molecular heterogeneity in SNX27–retromer-mediated endosome-to-plasma-membrane recycling. *J. Cell Sci.* 127 (22), 4940–4953. doi: 10.1242/jcs.156299

Miyaguchi, K. (2004). Localization of selenium-binding protein at the tips of rapidly extending protrusions. *Histochem. Cell Biol.* 121 (5), 371–376. doi: 10.1007/s00418-004-0623-y

Mollet, J., Delahodde, A., Serre, V., Chretien, D., Schlemmer, D., Lombes, A., et al. (2008). CABC1 Gene Mutations Cause Ubiquinone Deficiency with Cerebellar Ataxia and Seizures. *Am. J. Hum. Genet.* 82 (3), 623–630. doi: 10.1016/j.ajhg.2007.12.022

Monks, S. A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., et al. (2004). Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* 75 (6), 1094–1105. doi: 10.1086/426461

Pampouille, E., Berri, C., Boitard, S., Hennequet-Antier, C., Beauclercq, S. A., Godet, E., et al. (2018). Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. *BMC Genomics* 19 (1), 202. doi: 10.1186/s12864-018-4598-9

Peansukmanee, S., Vaughan-Thomas, A., Carter, S. D., Clegg, P. D., Taylor, S., Redmond, C., et al. (2009). Effects of hypoxia on glucose transport in primary equine chondrocytes *in vitro* and evidence of reduced GLUT1 gene expression in pathologic cartilage *in vivo*. *J. Orthopaedic Res.* 27 (4), 529–535. doi: 10.1002/jor.20772

Ponsuksili, S., Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M. (2015). Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. *PloS One* 10 (4), e0123678. doi: 10.1371/journal.pone.0123678

Putiri, E., and Pelegri, F. (2011). The zebrafish maternal-effect gene mission impossible encodes the DEAH-box helicase Dhx16 and is essential for the expression of downstream endodermal genes. *Dev. Biol.* 353 (2), 275–289. doi: 10.1016/j.ydbio.2011.03.001

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria:R Foundation for Statistical Computing.

Rederstorff, M., Krol, A., and Lescure, A. (2006). Understanding the importance of selenium and selenoproteins in muscle function. *Cell. Mol. Life Sci.* 63 (1), 52–59. doi: 10.1007/s00018-005-5313-y

Renault, A.-L., Mebirouk, N., Cavaciuti, E., Le Gal, D., Lecarpentier, J., d'Enghien, C. D., et al. (2017). Telomere length, ATM mutation status and cancer risk in Ataxia-Telangiectasia families. *Carcinogenesis* 38 (10), 994–1003. doi: 10.1093/carcin/bgx074

Rivero, J. L., Talmadge, R. J., and Edgerton, V. R. (1996). Correlation between myofibrillar ATPase activity and myosin heavy chain composition in equine skeletal muscle and the influence of training. *Anat. Rec.* 246 (2), 195–207. doi: 10.1002/(sici)1097-0185(199610)246:2<195::aid-ar6>3.0.co;2-0

Rivero, J. L., Sporleder, H. P., Quiroz-Rothe, E., Vervuert, I., Coenen, M., and Harmeyer, J. (2002). Oral L-carnitine combined with training promotes changes in skeletal muscle. *Equine Vet. J. Suppl.* (34), 269–274. doi: 10.1111/j.2042-3306.2002.tb05431.x

Rivero, J. L. (2004). Muscle morphology heterogeneity: control, significance for performance and responses to training, in *Emerging Equine Science*. Eds. Alliston, J. C., Ede, S., Hemmings, A., Hyslop, A., Longland, J., Moreton, A., and Moore-Colyer, H. (Nottingham:Nottingham University Press), 11–27. doi: 10.1017/S0263967X00041197

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Rooney, M. F., Porter, R. K., Katz, L. M., and Hill, E. W. (2017). Skeletal muscle mitochondrial bioenergetics and associations with myostatin genotypes in the Thoroughbred horse. *PloS One* 12 (11), e0186247. doi: 10.1371/journal.pone.0186247

Rudich, A., Konrad, D., Torok, D., Ben-Romano, R., Huang, C., Niu, W., et al. (2003). Indinavir uncovers different contributions of GLUT4 and GLUT1 towards glucose uptake in muscle and fat cells and tissues. *Diabetologia* 46 (5), 649–658. doi: 10.1007/s00125-003-1080-1

Sajuthi, S. P., Sharma, N. K., Chou, J. W., Palmer, N. D., McWilliams, D. R., Beal, J., et al. (2016). Mapping adipose and muscle tissue expression quantitative trait

loci in African Americans to identify genes for type 2 diabetes and obesity. *Hum. Genet.* 135 (8), 869–880. doi: 10.1007/s00439-016-1680-8

Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K., et al. (2017). Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* 18 (1), 565. doi: 10.1186/s12864-017-3943-8

Serrano, A. L., Quiroz-Rothe, E., and Rivero, J.-L. L. (2000). Early and long-term changes of equine skeletal muscle in response to endurance training and detraining. *Pflügers Archiv.* 441 (2), 263–274. doi: 10.1007/s004240000408

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis *via* large matrix operations. *Bioinformatics* 28 (10), 1353–1358. doi: 10.1093/bioinformatics/bts163

Sharma, N. K., Langberg, K. A., Mondal, A. K., Elbein, S. C., and Das, S. K. (2011). Type 2 diabetes (T2D) associated polymorphisms regulate expression of adjacent transcripts in transformed lymphocytes, adipose, and muscle from Caucasian and African-American subjects. *J. Clin. Endocrinol. Metab.* 96 (2), E394–E403. doi: 10.1210/jc.2010-1754

Snow, D. H., Harris, R. C., and Gash, S. P. (1985). Metabolic response of equine muscle to intermittent maximal exercise. *J. Appl. Physiol. (1985)* 58 (5), 1689–1697. doi: 10.1152/jappl.1985.58.5.1689

Snow, D.H.a.V. SJ. (1994). Muscle anatomy, physiology and adaptations to exercise and training in *The athletic horse: principles and practice of equine sports medicine*. Ed.Hodgson, D. R. R. (Philadelphia:Saunders), 145–179.

Stadler, S. C., Polanetz, R., Meier, S., Mayerhofer, P. U., Herrmann, J. M., Anslinger, K., et al. (2005). Mitochondrial targeting signals and mature peptides of 3-methylcrotonyl-CoA carboxylase. *Biochem. Biophys. Res. Commun.* 334 (3), 939–946. doi: 10.1016/j.bbrc.2005.06.190

Stadler, S. C., Polanetz, R., Maier, E. M., Heidenreich, S. C., Niederer, B., Mayerhofer, P. U., et al. (2006). Newborn screening for 3-methylcrotonyl-CoA carboxylase deficiency: population heterogeneity of MCCA and MCCB mutations and impact on risk assessment. *Hum. Mutat.* 27 (8), 748–759. doi: 10.1002/humu.20349

Stefely, J. A., and Pagliarini, D. J. (2017). Biochemistry of mitochondrial coenzyme Q biosynthesis. *Trends In Biochem. Sci.* 42 (10), 824–843. doi: 10.1016/j.tibs.2017.06.008

Stefely, J. A., Licitra, F., Laredj, L., Reidenbach, A. G., Kemmerer, Z. A., Grangeray, A., et al. (2016). Cerebellar ataxia and coenzyme Q deficiency through loss of unorthodox kinase activity. *Mol. Cell* 63 (4), 608–620. doi: 10.1016/j.molcel.2016.06.030

Stranger, B. E., Nica, A. C., Forrest, M. S., Dimas, A., Bird, C. P., Beazley, C., et al. (2007). Population genomics of human gene expression. *Nat. Genet.* 39 (10), 1217–1224. doi: 10.1038/ng2142

Taylor, R. M., and Sunde, R. A. (2017). Selenium requirements based on muscle and kidney selenoprotein enzyme activity and transcript expression in the turkey poult (Meleagris gallopavo). *PloS One* 12 (11), e0189001. doi: 10.1371/journal.pone.0189001

Timmons, J. A., Knudsen, S., Rankinen, T., Koch, L. G., Sarzynski, M., Jensen, T., et al. (2010). Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J. Appl. Physiol.* 108 (6), 1487–1496. doi: 10.1152/japplphysiol.01295.2009

Todd, E. T., Ho, S. Y. W., Thomson, P. C., Ang, R. A., Velie, B. D., and Hamilton, N. A. (2018). Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Sci. Rep.* 8 (1), 6167. doi: 10.1038/s41598-018-24663-x

Turunen, M., Olsson, J., and Dallner, G. (2004). Metabolism and function of coenzyme Q. *Biochim. Biophys. Acta (BBA) - Biomembr.* 1660 (1), 171–199. doi: 10.1016/j.bbamem.2003.11.012

Tyler, C. M., Golland, L. C., Evans, D. L., Hodgson, D. R., and Rose, R. J. (1998). Skeletal muscle adaptations to prolonged training, overtraining and detraining in horses. *Pflugers Arch.* 436 (3), 391–397. doi: 10.1007/s004240050648

Valette, J. P., Barrey, E., Jouglin, M., Courouce, A., Auvinet, B., and Flaux, B. (1999). Standardisation of muscular biopsy of gluteus medius in French trotters. *Equine Vet. J. Suppl.* (30), 342–344. doi: 10.1111/j.2042-3306.1999.tb05246.x

Velez-Irizarry, D., Casiro, S., Daza, K. R., Bates, R. O., Raney, N. E., and Steibel, J. P., et al. (2019). Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs. *BMC Genomics* 20 (1), 3. doi: 10.1186/s12864-018-5386-2

Votion, D. M., Gnaiger, E., Lemieux, H., Mouithys-Mickalad, A., and Serteyn, D. (2012). Physical Fitness and Mitochondrial Respiratory Capacity in Horse Skeletal Muscle. *PloS One* 7 (4). doi: 10.1371/journal.pone.0034890

Wang, G., Durussel, J., Shurlock, J., Mooses, M., Fuku, N., Bruinvels, G., et al. (2017). Validation of whole-blood transcriptome signature during microdose recombinant human erythropoietin (rHuEpo) administration. *BMC Genomics* 18 (Suppl 8), 817. doi: 10.1186/s12864-017-4191-7

Westra, H. J., and Franke, L. (2014). From genome to function by studying eQTLs. *Biochim. Biophys. Acta* 1842 (10), 1896–1902. doi: 10.1016/j.bbadis.2014.04.024

Wickham, H. (2009). *Ggplot2: elegant graphics for data analysis*. New York:Springer.

Willett, P. (1975). *An introduction to the thoroughbred*. London:Paul.

Wittkopp, P. J. (2005). Genomic sources of regulatory variation in cis and in trans. *Cell Mol. Life Sci.* 62 (16), 1779–1783. doi: 10.1007/s00018-005-5064-9

Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., et al. (2014). Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* 46 (5), 430–437. doi: 10.1038/ng.2951

Xia, Y., Brown, L., Yang, C. Y., Tsan, J. T., Siciliano, M. J., Espinosa, R., et al. (1991).TAL2, a helix-loop-helix gene activated by the formatting translocation in human T-cell leukemia. *Proc. Natl. Acad. Sci.* 88 (24), 11416–11420. doi: 10.1073/pnas.88.24.11416

Yang, M., and Sytkowski, A. J. (1998). Differential expression and androgen regulation of the human selenium-binding protein gene hSP56 in prostate cancer cells. *Cancer Res.* 58 (14), 3150–3153.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J. Integr. Biol.* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Zisman, A., Peroni, O. D., Abel, E. D., Michael, M. D., Mauvais-Jarvis, F., Lowell, B. B., et al. (2000). Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nat. Med.* 6 (8), 924–928. doi: 10.1038/78693

Conflict of Interest: EH was employed by Plusvital Ltd. Plusvital Ltd had no role, financial or otherwise, in the research.EH, DM and LK are named inventors on multiple international patents relating to the application of variation in the prediction of race distance performance; none of which is relevant to the data/results reported in this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Farries, Bryan, McGivney, McGettigan, Gough, Browne, MacHugh, Katz and Hill. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 January 2020
doi: 10.3389/fgene.2019.01263

Integrated Analysis of Methylome and Transcriptome Changes Reveals the Underlying Regulatory Signatures Driving Curly Wool Transformation in Chinese Zhongwei Goats

Ping Xiao[1,2], Tao Zhong[2], Zhanfa Liu[3], Yangyang Ding[1], Weijun Guan[1], Xiaohong He[1], Yabin Pu[1], Lin Jiang[1], Yuehui Ma[1] and Qianjun Zhao[1]**

[1] Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China, [3] The Ningxia Hui Autonomous Region Breeding Ground of Zhongwei Goat, Department of Agriculture and Rural Areas of Ningxia Hui Autonomous Region, Wuzhong, China

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
Yun Li,
Ocean University of China, China
Mairead Lesley Bermingham,
University of Edinburgh,
United Kingdom

***Correspondence:**
Yuehui Ma
yuehui.ma@263.net
Qianjun Zhao
zhaoqianjun@caas.cn

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 20 February 2019
Accepted: 15 November 2019
Published: 08 January 2020

Citation:
Xiao P, Zhong T, Liu Z, Ding Y, Guan W, He X, Pu Y, Jiang L, Ma Y and Zhao Q (2020) Integrated Analysis of Methylome and Transcriptome Changes Reveals the Underlying Regulatory Signatures Driving Curly Wool Transformation in Chinese Zhongwei Goats. Front. Genet. 10:1263. doi: 10.3389/fgene.2019.01263

The Zhongwei goat is kept primarily for its beautiful white, curly pelt that appears when the kid is approximately 1 month old; however, this representative phenotype often changes to a less curly phenotype during postnatal development in a process that may be mediated by multiple molecular signals. DNA methylation plays important roles in mammalian cellular processes and is essential for the initiation of hair follicle (HF) development. Here, we sought to investigate the effects of genome-wide DNA methylation by combining expression profiles of the underlying curly fleece dynamics. Genome-wide DNA methylation maps and transcriptomes of skin tissues collected from 45- to 108-day-old goats were used for whole-genome bisulfite sequencing (WGBS) and RNA sequencing, respectively. Between the two developmental stages, 1,250 of 3,379 differentially methylated regions (DMRs) were annotated in differentially methylated genes (DMGs), and these regions were mainly related to intercellular communication and the cytoskeleton. Integrated analysis of the methylome and transcriptome data led to the identification of 14 overlapping genes that encode crucial factors for wool fiber development through epigenetic mechanisms. Furthermore, a functional study using human hair inner root sheath cells (HHIRSCs) revealed that, one of the overlapping genes, platelet-derived growth factor C (*PDGFC*) had a significant effect on the messenger RNA expression of several key HF-related genes that promote cell migration and proliferation. Our study presents an unprecedented analysis that was used to explore the enigma of fleece morphological changes by combining methylome maps and transcriptional expression, and these data revealed stage-specific epigenetic changes that potentially affect fiber development. Furthermore, our functional study highlights a possible role for the overlapping gene *PDGFC* in HF cell growth, which may be a predictable biomarker for fur goat selection.

Keywords: Zhongwei goat, deoxyribonucleic acid methylation, curly pelts, epigenetics, transcriptomics platelet-derived growth factor C

INTRODUCTION

Animal hair fibers and fur are essential raw materials for the textile industry, and people have been taming and improving some fur- and wool-producing animals, such as sheep, goats, and rabbits. Unlike cashmere goats, Chinese Zhongwei goats have a reputation for their pelts, which have white, lustrous staples and attractive curls when they are obtained at approximately 35 days of age, with fibers on the skins of kids comprising 86% heterotypic fibers and 14% true wool by weight (Gong, 1994). Nevertheless, these natural and exquisite patterns are becoming less economically valuable as the curly form of the wool disappears within 2 months of the kid's life, and the exact reasons for its disappearance remain elusive.

A few studies have found that genetic polymorphisms in candidate genes can account for various hair traits in different species (Adhikari et al., 2016; Demars et al., 2017; Morgenthaler et al., 2017). Some critical signaling pathways, including the wingless-related integration site (WNT), ectodysplasin A receptor (EDAR), and bone morphogenetic proteins (BMP) pathways, are regarded as regulatory hubs during fiber development (Lu et al., 2016; Telerman et al., 2017). Mammals, and in particular, sheep and goats have obvious periodic fiber growth with seasonal changes, and the regulatory mechanisms of these changes have been explored in transcriptome studies (Yang et al., 2017; Li et al., 2018). Subsequently, some genetic factors, such as those of the *TCHH*, *KRT* gene families, and the *metallothionein 3 isoforms*, which are related to curly wool, have been determined by RNA-seq analysis and immunohistochemical analyses of the fiber proteins (Sriwiriyanont et al., 2011; Yu et al., 2011; Kang et al., 2013). These findings may thus underpin this dynamic morphogenesis.

The epigenome, which contains a great deal of modifiable genetic information, is the source of many determining factors in regulatory mechanisms. Increased DNA methylation and histone modification status may enhance pathological immune responses and suppress hair follicle (HF) development in anagen (Zhao et al., 2012). Diverse whole genome methylation profiles have been found to characterize the two periods (anagen and telogen) of HF growth, suggesting that increased transcript expression levels are connected with compromised DNA methylation (Bock et al., 2012). Highly expressed *DNA methyltransferase 1* (*DNMT1*) can prevent the epithelial progenitor cells in the HF bulb from overproliferating to drive differentiation, thus maintaining a normal HF structure (Sen et al., 2010). Because of the tight connection between individual development (including skin development, regeneration, and HF cycling) and DNA methylation (Gudjonsson and Krueger 2012; Botchkarev et al., 2013; Plikus et al., 2015), it is necessary to concentrate the genome-wide methylation profile of dynamic hair morphogenesis.

In the present study, we assessed DNA methylation profiles by whole genome bisulfite sequencing (WGBS) and transcriptional expression by RNA sequencing analysis (RNA-seq) of shoulder skin samples from 45-day-old kids with curly wool and from the same kids exhibiting non-curling wool at 108 days. Integrated analysis (WGBS and RNA-seq) selected differentially expressed-methylated gene (DEGs-DMGs) candidates, which tend to be the key factors in curly wool development through epigenetic patterns. Eventually, the promoting effect of the platelet-derived growth factor C (*PDGFC*) gene on HF cell growth was validated through a functional study *in vitro*. This study confirms the importance of an integrated analysis that combines DNA methylation and gene expression for determining curly hair traits and provides comprehensive resources for studying HF development in humans.

MATERIALS AND METHODS

Animals and Sample Preparation

Three Chinese Zhongwei goats bred at the breeding base of Zhongwei goats (located in the Ningxia Hui Autonomous Region, China) were randomly selected for this study. The goats had no kinship relevant to their use as samples, and they were raised under the same conditions to minimize external factors. When they were 45 and 108 days old, we cleaned their hair and disinfected the target in the scapular region, from which skin samples were collected using sterilized scalpel blades. Some samples were immediately stored in RNAlater (Thermo Fisher Scientific, USA) for storage at −80°C until further processing, and some samples were rapidly stored in a 4% paraformaldehyde fixation solution to prepare paraffin sections. All the resulting wounds were treated with Yunnan Baiyao powder (Yunnan Baiyao Group Co., Ltd., China) to stop the bleeding. All of the animal experimental procedures were performed in accordance with the guidelines for the care and use of experimental animals established by the Ministry of Agriculture of the People's Republic of China and approved by Institute of Animal Science, Chinese Academy of Agricultural Sciences.

Ribonucleic Acid Isolation and Sequencing

Samples taken at 45 days and 108 days, representing curly haired (D45) and wavy haired (D108) individuals, respectively, were stored separately in RNAlater. Total RNA was extracted from these six samples by RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. The RNA quantity and quality were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA), and the RNA Integrity Number value of these samples was determined to be greater than 7.5, which was important to ensure RNA integrity. The RNA library construction, quality control, and sequencing were conducted using an Illumina Nova seq platform at the Berry NGS Company (Beijing, China), through which approximately 55 million paired-end reads (2 × 150 bp) were produced for each of samples. Before the downstream analysis, filtration of sequencing reads was conducted to remove the reads containing joints and to eliminate the low-quality reads. The remaining clean data were matched to the reference genome at CHIR_1.0 (September 10, 2015, ftp://ftp.ncbi.nlm.nih.gov/

genomes/all/GCA/000/317/765/GCA_000317765.1_CHIR_1.0/) using the new version of HISAT2 (v2.1.0) (Kim et al., 2015). The average alignment rate for the RNA-seq was 83.53% (82.63–84.70%, median = 83.45%) (**Supplementary Table S1**), and then, the transcripts were assembled, quantified and merged with StringTie (v1.3.4). The output files were prepared for use in the differential expression analysis. RNA-seq data were deposited into a NCBI BioProject section under accession number PRJNA524985.

Differential Expression Analysis

Output files containing the expression levels of the exons, introns, and transcript of each sample were processed with the Ballgown R package (v2.14.1) (Frazee et al., 2015). A parametric F-test using "stattest" function in Ballgown module was used to compare transcript abundance, and the covariate "curly (samples in 45 days)/wave (samples in 108 days)" was corrected for the calculated p value. The differential expressed transcripts (DETs) were annotated to gene names to identify and list the DEGs, which were screened based on the following criteria: p value ≤ 0.05 and absolute fold change value > 1.5. The heatmap of DEG hierarchical clustering was performed using the R package pheatmap (v1.0.10).

Differentially Expressed Genes Profile Analysis

We selected eight DEGs randomly and validated their expressions by using quantitative (qPCR) on an ABI 7500 (Applied Biosystems, USA) in a 20-μl reaction containing 2 μl of the complementary DNA template (generated by the reverse transcription kit, Takara), 10 μl of 2 × SYBR Green Master Mix (RR420A, Takara), and 0.8 μl of each primer (10 μmol/μl), with glyceraldehyde 3-phosphate dehydrogenase as the endogenous control. The primers used for qPCR were designed with Primer Premier 5 (v5.00, http://www.premierbiosoft.com) and are listed in **Supplementary Table S2**. The transcription factor binding site analysis was performed by uploading our DEG list into the Innate DB database (v5.3) (Breuer et al., 2013), where the data were subjected to the hypergeometric algorithm and with Benjamini-Hochberg correlation method (p value ≤ 0.05). A transcript splicing analysis of the DEGs was carried out by combining the extracted splice-sites information from the genome annotation results and our RNA-seq data, and we compared the alternative splicing events of each gene in two periods. Differential alternative splicing events, including skipped exon (SE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), mutually exclusive exon (MXE), and retained intron (RI) events were detected using rMATS (v 4.0.1) (Shen et al., 2014), and the events were considered as significantly different based on the following filtering criteria: | Inc level difference | ≥ 5% and false discovery rate (FDR) < 0.01. To explore the potential relationships among expressed DEGs, a weighted correlation network analysis (WGCNA) was performed using the WGCNA R package (v 1.68) (Langfelder and Horvath, 2008) with 326 DEGs used as input data. We retained the genes ranked in the top 90% of the variance size between two groups, and retained 242 DEGs to generate correlation networks. The soft threshold value was set as 8. The correlation between eigenvectors of each module and the curl-wavy status was calculated, and those with |coefficient value| > 0.8 and p value < 0.05 were considered to significant.

Whole Genome Bisulfite Sequencing

Genomic DNA (from three samples taken at 45 days, three samples taken at 108 days) were isolated from scapular skin tissues using Wizard Genomic DNA purification kit (Promega, Madison, USA) following the manufacturer's instructions. Constructed DNA libraries were sequenced at the Berry NGS Company (Beijing, China) using an Illumina Nova seq 6000 platform (Illumina, San Diego, CA, USA), and the subsequent raw reads were filtered to remove contaminated reads in three steps: 1) removing any read that contained a 3' adapter oligonucleotide sequence, 2) removing any read for which the percentage of Ns (unknown bases) was > 10%, and 3) removing any low quality reads (Phred score ≤ 5, percentage of low quality bases ≥ 50%). Then, an average of 600,000,000 paired-end 150-bp reads was acquired for the six samples. Next, lambda sequences were included in the clean reads to evaluate the C-T conversion rate. The sample information for methylation sequencing data were submitted to the NCBI BioProject section under accession number PRJNA555706.

Deoxyribonucleic Acid Methylation Data Analysis

The quality controlled clean reads were converted into bisulfite-treated status reads (C-to-T and G-to-A transformed) before being aligned with the corresponding bisulfite-converted goat reference genome, CHIR_1.0 (Dong et al., 2013), using Bismark (v 0.7.12) (Krueger and Andrews, 2011). After the reads were processed as BAM files by using SAMtools (v 1.9) (Li et al., 2009), Picard software (v 1.96) was used to remove duplicate reads, and the data on methylation status per cytosine site were extracted based on the Bismark instruction manual. Methylation can be identified by determining the methylation level of a specific cytosine site according to the following formula: averaged methylation level = counts of methylated reads/counts of total reads × 100%. We then used the scriptlet "bismark_methylation_extractor" in Bismark to analyze the methylation status throughout the whole genomic DNA and the methylation distribution in various genomic regions (including upstream, downstream, gene, exonic, intronic, and intergenic regions). We identified the differentially methylated regions (DMRs, with a 500 bp window) based on the sliding window approach combined with a logistic regression method using methylKit software (v 1.10.0) to compare the methylation status of specific regions in the two groups (Akalin et al., 2012). The following screening criteria were used for obtaining the DMRs: 1) the average methylation difference between two pairwise groups was > 0.25; 2) the FDR value of methylation difference was <0.05; and 3) all DMRs were in uniquely mapped regions. The obtained DMRs were annotated to different genomic regions, where they could be considered DMGs because they had overlapping methylated cytosine (mC) sites in functional gene regions.

Polymerase Chain Reaction Validation of the Bisulfite Sequencing

A total of 500 ng of genomic DNA previously extracted from each sample at the same time points (three samples) were mixed together and treated with bisulfite with an EZ DNA Methylation-Gold Kit (Zymo Research). The information on the primers used for bisulfite sequencing PCR is listed in **Supplementary Table S3**. Bisulfite-treated products were amplified by High Fidelity Taq DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions. The PCR products purified by DNA Gel Extraction Kit (Qiagen) were ligated to a T-vector plasmid (TransGen Biotech), and the plasmids were transformed into *Escherichia* coli DH5α competent cells (Takara). We selected 10 single amplified clones for each group after they had incubated on a solid medium. Sequencing detection for the single clones was conducted by Sangon Biotech (Shanghai). Sequence data were analyzed by the online DNA methylation analysis platform at the BISMA website (http://services.ibc.uni-stuttgart.de/BDPC/BISMA/).

Gene Functional Enrichment Analysis

We used the g:Profiler web server (v 0.6.7) (Reimand et al., 2016) to conduct the Gene Ontology (GO) enrichment analysis of the DEGs and DMGs. All genes with average expression FPKM (fragments per kilobase of transcript per million fragments mapped) > 1 from all the samples were used as the background gene set for this analysis, and GO terms in which the P value ≤ 0.05, as corrected by the g:SCS threshold method (a significance criterion in g:Profiler), were considered significant. An analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment for the DEGs and DMGs was conducted through the online software KOBAS (v3.0) (Wu et al., 2006) to detect the related signaling pathway of each candidate gene set (BH-corrected P value < 0.05).

Protein Interaction Network of Integrated Genes

We used the STRING database (v10.5) (Szklarczyk et al., 2017) to construct and screen for a protein–protein interaction (PPI) network that contained differentially methylated and expressed genes (D45 versus D108). We only retained edges of the network that meet the following parameters: confidence score >0.8 and combined score > 0.8. Cytoscape (v3.6.0) (http://www.cytoscape.org/) was used to visualize interactions for the gene-gene pair input, including their combined score and the expression and methylation trends.

Cell Culture and Downstream Validation

Since the functional effects on fiber shape are changed, we used human hair inner root sheath cells (HHIRSCs) for further validation. We obtained HHIRSCs from ScienCell research laboratories (Carlsbad, CA, USA). The cells were incubated in mesenchymal cell medium (ScienCell Research Laboratories), which contained 1% mesenchymal stem cell growth supplement (MSCGS) (ScienCell Research Laboratories), 5% fetal bovine serum (FBS) (ScienCell Research Laboratories), 100 U/ml penicillin (ScienCell Research Laboratories), and 100 μg/ml streptomycin (ScienCell Research Laboratories) in a humidified, 37°C, 5% CO2 atmosphere, and the culture vessels were prepared with poly-L-lysine (2 μg/cm2) 1 day seeding to promote cell adherence. The cells were passaged through 4, but less than 5, population doublings to ensure their mesenchymal cell morphology and for use in other transfection experiments. After reviewing previous studies, we selected SMAD3 and *PDGFC* as candidate genes because they are potentially epigenetically regulated and act on HFs or on epidermal cell development during initial follicle formation. We conducted our pre-experiment using mouse fibroblasts (NIH/3T3 cells) to determine whether the overexpression of the selected genes had effects on the key signatures involved in the development of HFs (data not shown). We then chose *PDGFC* as the candidate to use in our further validation experiments with the human inner root sheath cells (HHIRSCs) since only *PDGFC* had a significant effect on the signatures of HFs development. The full-length coding DNA Sequence (CDS) of Homo sapiens *PDGFC* was ligated to pIRES2-EGFP to construct the overexpression plasmid, while the negative control (pIRES2-EGFP-NC) was constructed without target genes. To insert the targeted gene into HHIRSC cells effectively, we used a Lipofectamine® 3000 Transfection Kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. Three small interfering RNA (siRNA) sequences (RiboBio, Guangzhou, China) were used to knock down the expression of *PDGFC* in the HHIRSCs: si-h-*PDGFC*_001, CCAACCTGAGTAGTAAATT; si-h-*PDGFC*_002, GGAACAGA ACGGAGTACAA; and si-h-*PDGFC*_003, GGAAGACCTTATTCGATAT. The siRNA transfection was conducted using a specific siRNA transfection kit and riboFECT™ CP reagent (RiboBio, Guangzhou, China).

All the processed cells were collected from six-well plates after 48 h of incubation for the RNA extraction using the TRIzol method. The qPCR of the related genes was performed according to methods described above, and the primer information is listed in **Supplementary Table S2**.

For evaluating the cell motility after we conducted different treatments, we performed a monolayer wound healing assay when cells were approaching 100% confluence, a wound was made by scratching the monolayer with a pipette tip, and then, the cells were incubated in the same condition as described. After 12 h, we compared the gaps with respect to the wound line among the different groups using program ImageJ software (v 1.52a) (NIH, Bethesda, MD, USA) and calculated the migration rate by the following equation: migration rate% = [1 − (wound gap at 12 h/wound gap at T0)] × 100%, where T0 represents the initial evaluation time, which was recorded immediately after the scratch was made.

Cell proliferation at four time points after the respective cell treatment (12, 18, 24, and 36 h) was evaluated using a Cell Counting Kit (CCK)-8 assay. HHIRSCs were seeded into 96-well plates at the same density (5×103 cells per well) before transfection, and each condition was replicated in four wells independently. After adding 10 μl of CCK-8 solution (Dojindo, Kumamoto, Japan) to each well for a 2 h incubation at 37°C, the

absorbance values at 450 nm were measured using a multifunctional spectrophotometer (Tecan Infinite 200 PRO, Tecan Group LTD, Austria).

RESULTS

To investigate the underlying mechanisms of wool fiber development in the two postnatal stages (45 and 108 days), representing curly and wavy fleece, respectively, we sampled scapular skin tissues from three unrelated Chinese Zhongwei goats at these two time point (**Figure 1**). Furthermore, we detected the HF structure at these two time points, and we found that the HF taken at 108 days had a more compact arrangement than that examined at 45 days (**Figures 1A**, **B**). We isolated DNA and RNA for further analysis by RNA-Seq and WGBS. After quality control and data refining, we obtained 22,034 gene transcripts and approximately 300 million methylation sites, which were included in our subsequent downstream analysis.

The Messenger Ribonucleic Acid Transcriptome Reveals Distinct Signatures in Dynamic Skin Development

The transcriptomic profiles obtained using RNA-seq were based on 41–42 and 38–50 million clean read pairs from the D45 and D108 samples, respectively and were uniquely mapped onto the *Capra hircus* CHIR_1.0 genome. All samples had at least 90.77% reads equal to or exceeding Q30 (**Supplementary Table S1**). These transcripts of the six samples showed a similar expression trend, with an average log2 (FPKM+1) value of approximately 2, suggesting the reliability of the general expression profiles (**Figure 2A**). For all the expression profiles, we found 326 DEGs, including 186 upregulated genes and 140 downregulated genes. We then verified the differential expression levels obtained from RNA-seq using qPCR, and the eight DEGs that were selected randomly followed the same trend as that of the sequencing data (**Figure 2B**). A hierarchical clustered map revealed that the expression of these 326 DEGs coincided with the same development period (**Figure 2C**).

We further explored whether the developmental process of HF affected alternative splicing events. We determined the distribution of the splicing events among the two groups and found no significant differences (**Supplementary Figure S1**), while the DEGs had a higher number of SE events (938, FDR < 0.01, Inc level difference > 5%) that were significantly different between the two groups (**Figure 3A**). A transcription factor (TF) enrichment analysis was performed to investigate the regulatory networks among these DEGs connected by coreactive TFs. For example, transcription factor *GLI family zinc finger 1 (GLI1)* and *interferon regulatory factor 2 (IRF2)* were significantly enriched in the promoters of upregulated and downregulated genes, respectively (**Figure 3B**).

To better understand the functions of these DEGs, a GO enrichment analysis was performed. The top five GO terms were cytoplasmic part (GO: 0044444), cytoplasm (GO: 0005737),

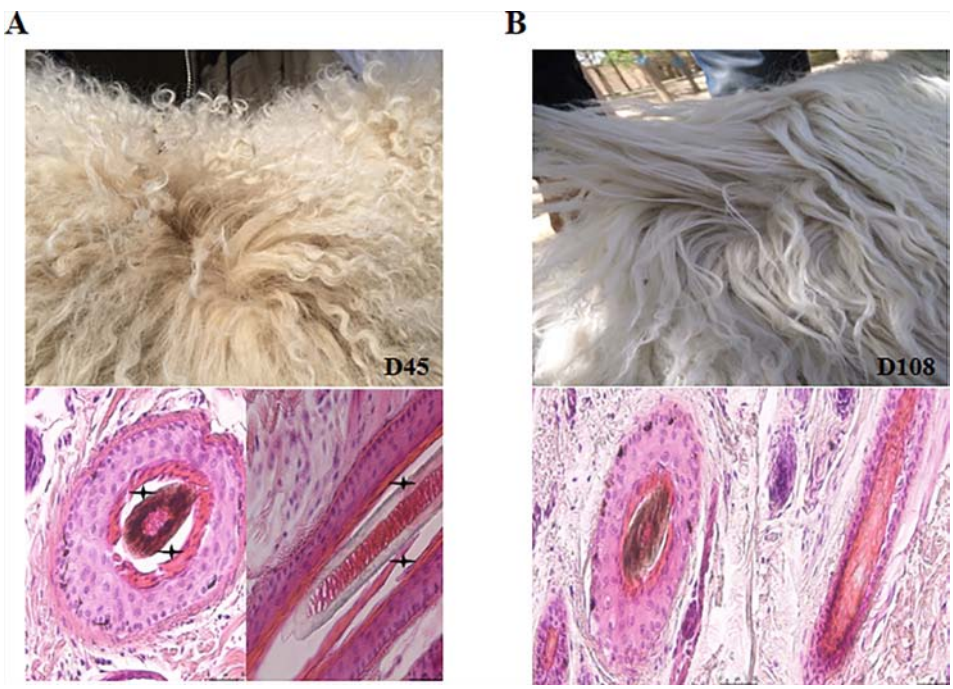


FIGURE 1 | The dynamics of fleece shape. **(A)** The fur of the Zhongwei goat at 45 days old and transverse/longitudinal sections of single hair follicle. **(B)** The fur of Zhongwei goat at 108 days old and transverse/longitudinal sections of a single hair follicle. Asterisk symbol highlights the gap between the hair shaft and inner root sheath at 45 days; the bar in all images of HE stained sections is 50 μm.

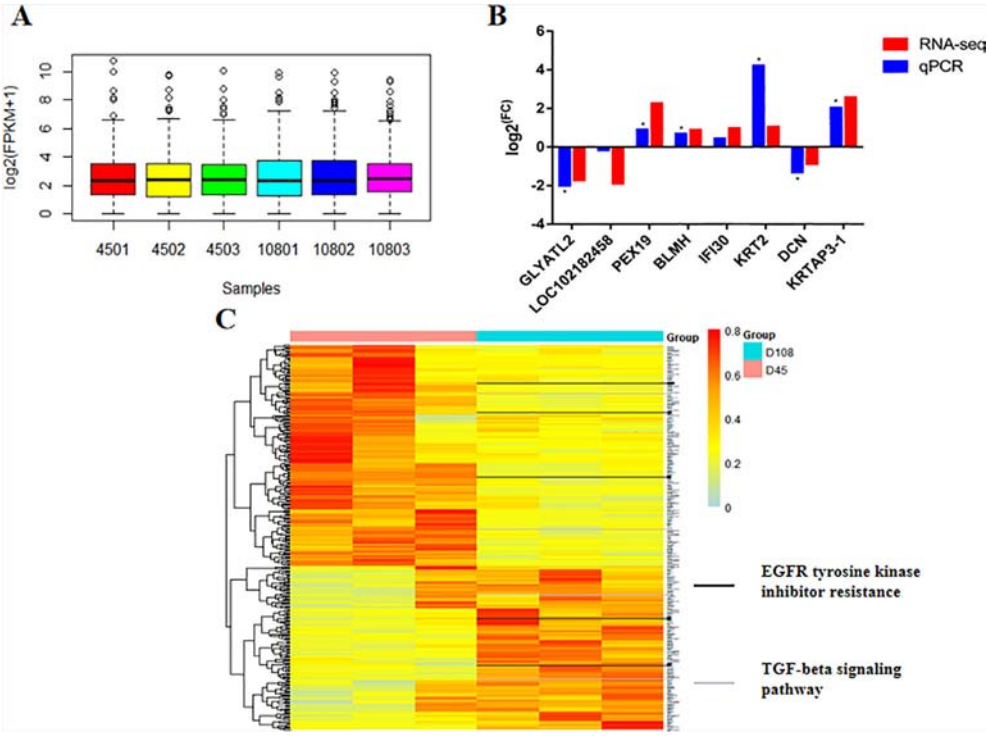


FIGURE 2 | The transcriptional profile of two groups during early hair growth. **(A)** Averaged RNA expression level among six individuals. **(B)** Validation of selected differently expressed genes (DEGs) through real-time quantitative PCR. The value is presented as the logarithmic form of the fold change (FC) between two groups. *P* values were calculated using Student's t tests (**P* < 0.05). **(C)** Clustered heatmap representation of DEG expression. Clustering was performed according to Pearson's correlation values. The black and gray bars outlined in the picture represent the DEGs that are involved in epidermal growth factor receptor tyrosine kinase inhibitor resistance and transforming growth factor beta signaling pathways, respectively.

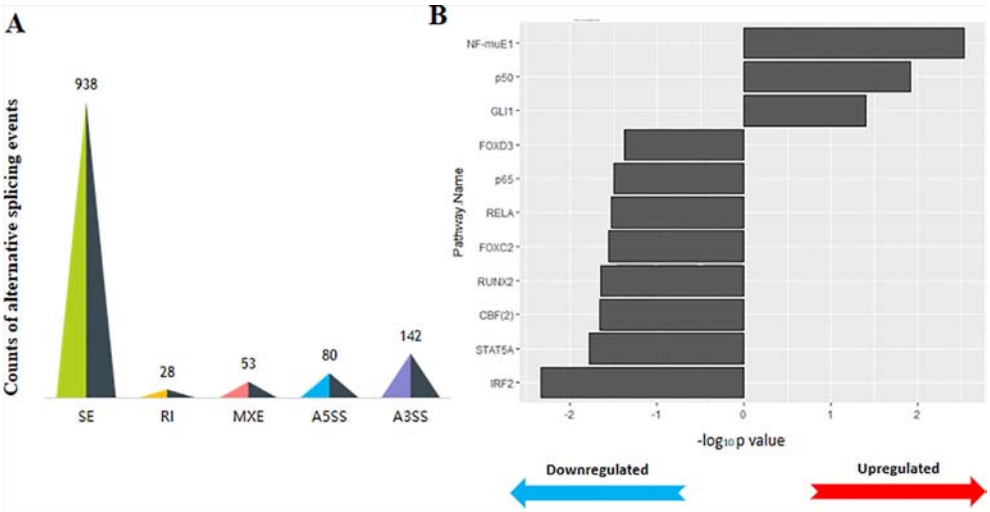


FIGURE 3 | The transcriptional process is changed during wool fiber development. **(A)** Count of the differential splicing events between D45 and D108 transcripts (*P* < 0.05). SE, skipped exon; RI, retained intron; MXE, mutually exclusive exons; A5SS, alternative 5' splice site; A3SS, alternative 3' splice site. There are 938 significantly different skipped exon events between the two stages, which was much greater than other events. **(B)** Transcription factor binding sites enriched among differentially expressed genes in D45 and D108. Only the transcription factors of upregulated or downregulated genes that had a *P* value < 0.01 were retained. All *P* values were corrected using the Benjamini–Hochberg method (BH-corrected *P* value < 0.05).

cytosol (GO: 0005829), skin development (GO: 0043588), and cornification (GO: 0070268) (**Supplementary Table S4**), which highlight the central roles of the cell conformation and keratinization in the DEGs during hair structure transformation. In the KEGG pathway enrichment, the antigen processing and presentation (chx04612), systemic lupus erythematosus (chx05322), epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor resistance (chx01521), fatty acid elongation (chx00062), asthma (chx05310), inflammatory bowel disease (chx05321), leishmaniasis (chx05140), selenocompound metabolism (chx00450), transforming growth factor beta (TGF)-beta signaling pathway (chx04350), and FoxO signaling pathway (chx04068) were considered significantly enriched pathways (**Table 1**).

Based on the WGCNA, we derived associated DEG expression coherence sets. There were 242 DEGs remaining after the expression comparison, blue (|coefficient of correlation| = 1, p-value = 2×10^{-5}) and turquoise (|coefficient of correlation| = 0.82, p-value = 5×10^{-2}) modules were correlated with curly wave status. The results indicated 91 and 151 genes were in the blue and turquoise modules, respectively (**Supplementary Figure S2**). We considered that a weight of gene-gene edges greater than 0.4 indicated a stable correlation, which led to the blue module having 255 edges and the turquoise module having 174 edges. A KEGG pathway enrichment analysis was also performed to determine the gain-the-function assessment of the DEGs in the two modules. Both modules were significantly enriched in EGFR tyrosine kinase inhibitor resistance and the PI3K-Akt signaling pathway (data not shown).

The Deoxyribonucleic Acid Methylation Profile Potentially Affects the Dynamic Transformation of Wool Fibers

Bisulfite sequencing enabled the acquisition of the genome-wide DNA methylation landscapes at single-base resolution of the postnatal D45 and D108 skin samples from Zhongwei goats. We obtained 360–459 million uniquely mapped reads among all the samples to ensure concordant coverage. The average ratio of uniquely mapped reads was 73.38% (71.30–74.65%, median = 73.66%), and the sequencing depths were all greater than 16 (**Table 2**).

The methylation level was calculated with the average of 321,036,397 and 322,677,466 methylated cytosines (mCs) in the 45- and 108-day stages, respectively. Among these detected cytosine sites, CpG (CG sites), as one in a nucleotide context, made up the highest proportion (84.71–85.34%), while

TABLE 1 | Kyoto Encyclopedia of Genes and Genomes pathway enrichment of differentially methylated genes and differently expressed genes.

KEGG pathway of DMGs (IDs)	Number of DMGs involved	P value	Overlapped genes	KEGG pathway of DEGs (IDs)	Number of DEGs involved	P value
Glutamatergic synapse (chx04724)	21	8.98054E-07		Antigen processing and presentation (chx04612)	6	0.001234391
Adherens junction (chx04520)	15	2.05829E-05		Systemic lupus erythematosus (chx05322)	7	0.004038687
Inflammatory mediator regulation of TRP channel (chx04750)	17	5.60775E-05		EGFR tyrosine kinase inhibitor resistance (chx01521)	5	0.007151845
Axon guidance (chx04360)	23	8.68786E-05		Fatty acid elongation (chx00062)	3	0.008275261
Long-term depression (chx04730)	12	9.18081E-05	SMAD3, *PDGFC*	Asthma (chx05310)	3	0.01251151
Oxytocin signaling pathway (chx04921)	21	9.7458E-05		Inflammatory bowel disease (chx05321)	4	0.016331757
Arrhythmogenic right ventricular cardiomyopathy (chx05412)	13	0.000158767		Leishmaniasis (chx05140)	4	0.023108939
Aldosterone synthesis and secretion (chx04925)	14	0.000169753		Selenocompound metabolism (chx00450)	2	0.026689839
Rap1 signaling pathway (chx04015)	23	0.000689252		TGF-beta signaling pathway (chx04350)	4	0.029163353
Gap junction (chx04540)	12	0.002506112		FoxO signaling pathway (chx04068)	5	0.037441899

Overlapped genes, the genes that are enriched in significant pathways as both DMGs and DEGs.

TABLE 2 | Summary of the whole-genome bisulfite sequencing dataset.

Sample	Clean reads	Unique mapped reads	Unique mapped ratio (%)	Average depth	Conversion rate (%)	%≥Q30
D4501	603,524,580	442,523,036	73.32	19.86	98.84	91.97; 88.77
D4502	598,727,366	442,986,932	73.99	18.93	98.78	91.96; 87.98
D4503	601,082,668	445,831,838	74.17	20.62	98.87	91.96; 88.79
D10801	505,312,936	360,303,746	71.30	16.63	98.73	92.01; 88.50
D10802	604,767,486	440,587,260	72.85	20.00	98.86	91.83; 89.14
D10803	614,904,624	459,049,754	74.65	23.03	98.91	91.77; 88.92

chlorhexidine gluconate (CHG) accounted for the smallest (3.47–3.58%) (**Figure 4A**). In addition to the broadest methylation distribution, these CpG sites had the highest average methylation level (72.26–72.69%), compared with the dramatically low methylation statuses found for the CHG (where H can be A, T or C) and CHH sites (0.56–0.59%) (**Figure 4B**). Although there was no obvious difference in methylation level between the two stages, we observed that, in distinct genomic features (downstream and upstream of the genes, exons, genes, intergenic regions, and introns), the regions upstream of the genes were weakly methylated at both CpG (< 50%) and non-CpG (< 0.6%) sites in all the samples, while exonic regions were the most methylated sites (CpG > 75% and non-CpG> 0.7%) (**Figure 4C**).

To identify the DMRs, we first computed the methylation status by analyzing the 500 bp-long sliding windows using the output of a methylKit (Akalin et al., 2012). A total of 3,379 DMRs were identified, including 1,651 hypermethylated DMRs and 2,128 hypomethylated DMRs in the 45-day sample (**Supplement Table S5**). A Manhattan plot was generated to show the DMR distribution along 30 chromosomes as –log10 (P-values) for all sliding windows (**Figure 5A**). The number of DMRs was reduced along with the number of chromosomes, but the DMR distribution was not affected, and a high density of DMRs was detected on chromosomes 13, 17, and 19. There were 1,471 DMRs annotated by gene name based on CHIR 1.0 assembly identification (Dong et al., 2013). We obtained 1,250 DMGs after merging data of the DMRs in the same gene, which contained 108 DMRs that were located in gene promoter regions [we considered areas upstream of the transcriptional start site (TSS) within 2,000 bp and downstream of the TSS within 200 bp as promoter regions]. Among all these DMGs, 635 were hypermethylated DMRs and 836 were hypomethylated DMRs, and 1,155 DMGs were annotated within intronic regions (**Figure 5B**). To examine the stability of the obtained DMRs, we randomly selected four DMRs that had region annotations (two intergenic, one exonic, and one upstream) of gene positions to validate the DNA methylation level. Bisulfite sequencing PCR (BSP) was used to detect the DNA methylated sites, although there was no significant difference in methylation between the samples taken at different time points, they presented similar trends in terms of methylation changes compared to those identified by the WGBS data (**Supplementary Figure S3**).

To explore the potential relationship of GO terms with these DMGs, a GO enrichment analysis was conducted by dividing these DMGs into hypermethylated and hypomethylated groups (**Supplementary Figure S4**). We identified the top 10 terms in three areas (cellular component, CC; biological process, BP; molecular function, MF), with specific descriptions, such as cell junction (GO: 0030054), cytoskeletal protein binding (GO: 0008092), cell development (GO: 0048468), cytoskeletal protein binding (GO: 0008092), and channel activity (GO: 0015267). For the KEGG pathway enrichment analysis performed on KOBAS

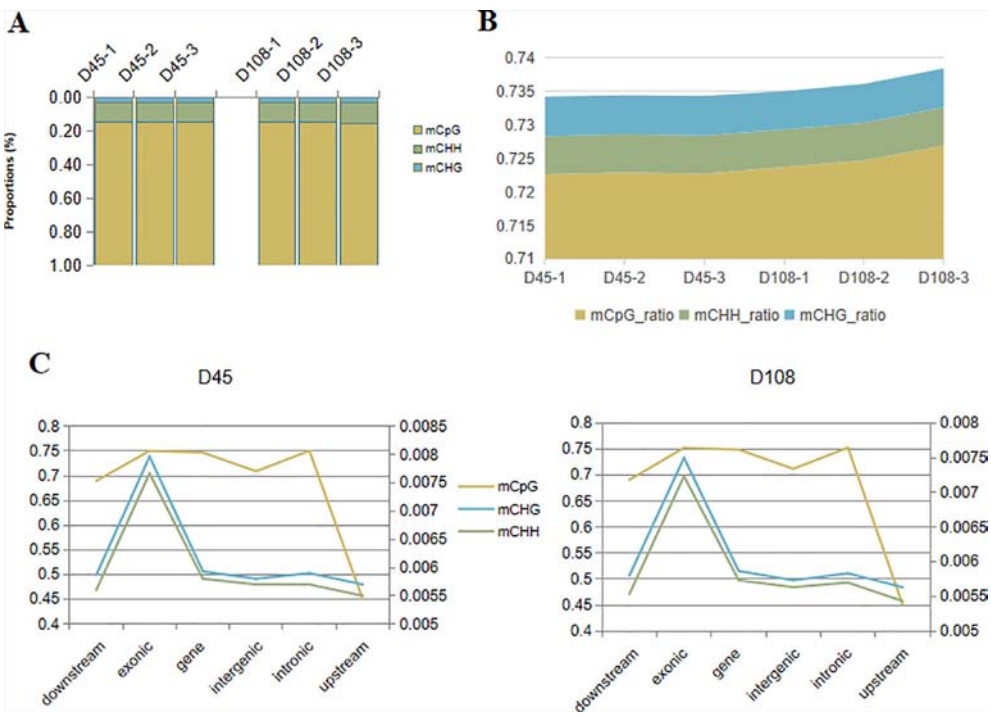


FIGURE 4 | DNA methylation profiles of skin samples in shift stages in the wool shape of Zhongwei goats. **(A)** The proportion of methylated cytosines (mCs) (mCpG, mCHH, and mCHG) in the D45 and D108 tissue samples. **(B)** The average methylation level (%) of cytosine sites [CpG, CHH, and chlorhexidine gluconate (CHG)] in six individual samples. **(C)** The methylation level (%) of three mCs in different genomic regions or elements. The vertical axis on the left represents the methylation levels of CpG in two stages, and the values on the right axis represents methylation levels of CHG and CHH.

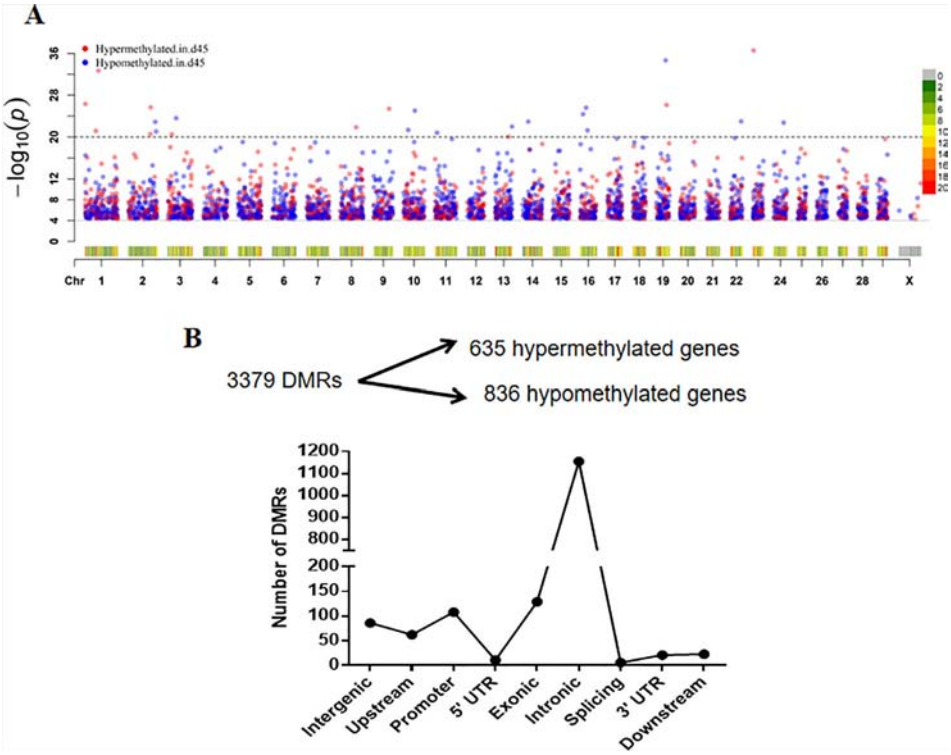


FIGURE 5 | Distribution of differentially methylated regions (DMRs) in different categories. **(A)** Manhattan plot of DMRs in a chromosomal landscape. Dots above the dotted line presented DMRs with –log10 (*p*) > 20. The heatmap below the dots represents the density (counts) of the DMR distribution within chromosomes. The blue and red dots indicate the status of hypomethylation and hypermethylation in the D45 regions, respectively. **(B)** The filtration of differentially methylated genes and the number of DMRs in different genomic regions.

v.3.0, we regarded adherens junction (chx04520) and gap junction (chx04540) as vital discoveries that illustrated the altered intercellular communication due to changes in epigenetic modification (**Table 1**).

An Integrated Analysis of the Differently Expressed Genes and Differentially Methylated Genes Was Used to Identify Candidate Genes That Control Hair Morphogenesis

To gain deeper insight into the RNA expression and DNA methylation differences linked to hair shaft development, we identified 14 overlapping genes among the 1,250 DMGs and 326 DEGS (**Figure 6A**). In these 14 overlapping genes, 9 annotated genes had differential methylation in the intronic regions, 4 genes were differentially methylated in intergenic areas (*SMAD3, CCDC91, MAP2,* and *SIK3*), and only 1, *LGMN,* had a DMR in the 3' UTR. We used 1,250 DMGs and their DNA methylation and gene expression data to explore their potential correlation. We found that 31 DMGs had a negative relationship (red and purple dots in **Figure 6B**), while 29 DMGs were positively regulated (lime green and blue dots in **Figure 6B**), and the DNA methylation status of 7 overlapping genes were associated with negatively regulated expression (*THADA, NOD1, MAP2, BLMH, LGMN, SMAD3,* and *SIK3*). Moreover, we conducted KEGG pathway enrichment to uncover the signaling pathway of DEGs and DMGs, and two genes, *SMAD3* and *PDGFC*, were involved in both of the most significantly enriched pathways of the DEGs and DMGs (**Table 1**). Then, we introduced a protein interactional network into these 14 overlapping genes to explore the mutual effects on the proteomics of hair developmental processes. Overall, four genes were found to apply to the scoring criterion that we set to ensure a strong association (confidence score >0.8 and combined score > 0.8) (**Figure 6C**). Two clusters are shown in this network analysis, which has had three integrated genes *(PDGFC, SMAD3,* and *NOD1*) in one network, and *DICER1* consists of an independent cluster with other imputed associated genes. The pathway analysis of the interactive network was performed, and 14 significant pathways were found, and we noticed some canonical signaling pathways, such as the TGF-beta signaling pathway, Hippo signaling pathway and FoxO signaling pathway (**Supplementary Table S6**).

PDGFC Is Associated With the Inner Root Sheath Cell Mesenchymal Phenotype and Enhances Hair Follicle Formation

To explore the role of *PDGFC* in the dynamic nature of HFs, cell culture, and overexpression, siRNA transfection was conducted using HHIRSCs. A wound healing assay showed that the

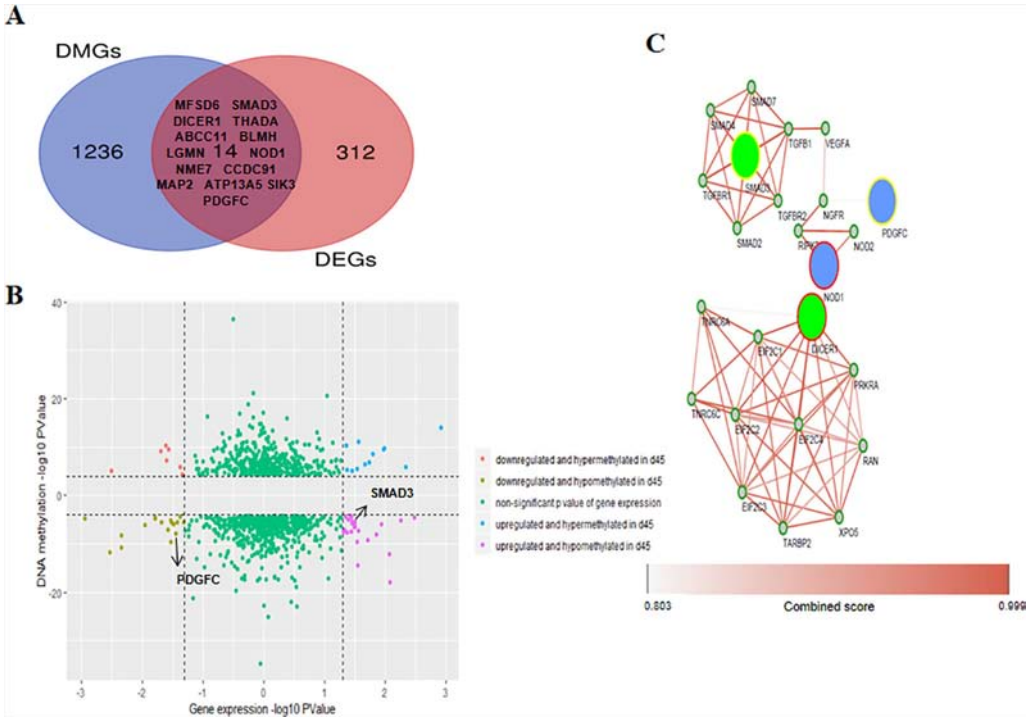


FIGURE 6 | Integrated analysis was used to identify genes with coupled differential DNA methylation and RNA transcription. **(A)** Venn diagram representing methylation-modified differently expressed genes during wool transformation. **(B)** Quadrant plot showing differentially methylated genes and expression levels of the corresponding genes. The vertical dotted lines indicate a threshold of the *P* value below 0.05, and parallel dotted lines show a threshold of the *P* value below 10^{-4}. **(C)** The protein-protein interaction (PPI) network among overlapping genes. Four genes identified from our analysis are included because both the confidence score and combined score were greater than 0.8. Circles filled with blue and green indicate downregulated and upregulated genes in D45, respectively. Circles with yellow and red margins indicate hypomethylation and hypermethylation in D45, respectively. Smaller gray circles with green margins denote genes attributed to the interaction.

HHIRSCs containing enhanced green fluorescent protein (EGFP)-*PDGFC* vectors had significantly stronger migration ability than the control group cells (EGFP) (**Figure 7A**). Correspondingly, cells with a specific RNA interference sequence of *PDGFC* exhibited relatively smaller migration areas compared with those of the group transfected with random small fragment sequences (**Figure 7A**). The cell proliferation rate was measured using CCK-8 assays, and groups with lower *PDGFC* expression (EGFP and si-*PDGFC*) showed a slower proliferation ratio at four selected time points compared with the case groups (**Figure 7B**). Furthermore, key genes associated with HF activation and development, for example, *GJA1* and *JAK1,* were upregulated and downregulated after *PDGFC* overexpressing gene transfection (**Figure 7C**). Notably, the mesenchymal marker *Vimentin* was upregulated in HHIRSC to maintain the morphologic characteristics of IRS. Not surprisingly, the suppression of *PDGFC* presented an opposite result *versus* the overexpression group (**Figure 7C**).

DISCUSSION

While previous studies have determined the effects of selected aspects, such as the genome-wide locus, DNA methylation, transcriptional signatures, and morphology, on the morphogenesis of curly fibers (Hynd et al., 2009; Cheng et al., 2010; Sriwiriyanont et al., 2011; Espada et al., 2014; Fan et al., 2015; Sennett et al., 2015; Gao et al., 2016; Glover et al., 2017; Li et al., 2018; Petridis et al., 2018), a comprehensive study conducting the interplay of genome-wide DNA methylation and transcription synchronously using skin tissues from the same biological replicates in different growth stages had thus far been lacking. The functional regulation of DNA methylation on gene expression has now been established as an effective prospective way to understand the ways that drastic methylation changes relate to hair phenotypic variation (Guo et al., 2014).

Overall, we observed 326 DEGs, some involved in EGFR tyrosine kinase inhibitor resistance, and the TGF-beta signaling pathway had been highlighted as valuable candidates for regulating HF development, based on previous studies (**Figure 2C**) (Cheng et al., 2010; Rognoni et al., 2014; Glover et al., 2017; Tripurani et al., 2018). Evidence from the transcription factor (TF) binding site analysis of the DEGs revealed the several TFs combine with the promoters of selected genes. For example, *interferon regulatory factor 2* (*IRF2*) and *signal transducer and activator of transcription 5A (STAT5A)* are reported TFs that serve as mediators to regulate transcriptional processes in HF growth and skin disease (Nishio et al., 2001; Legrand et al., 2016).

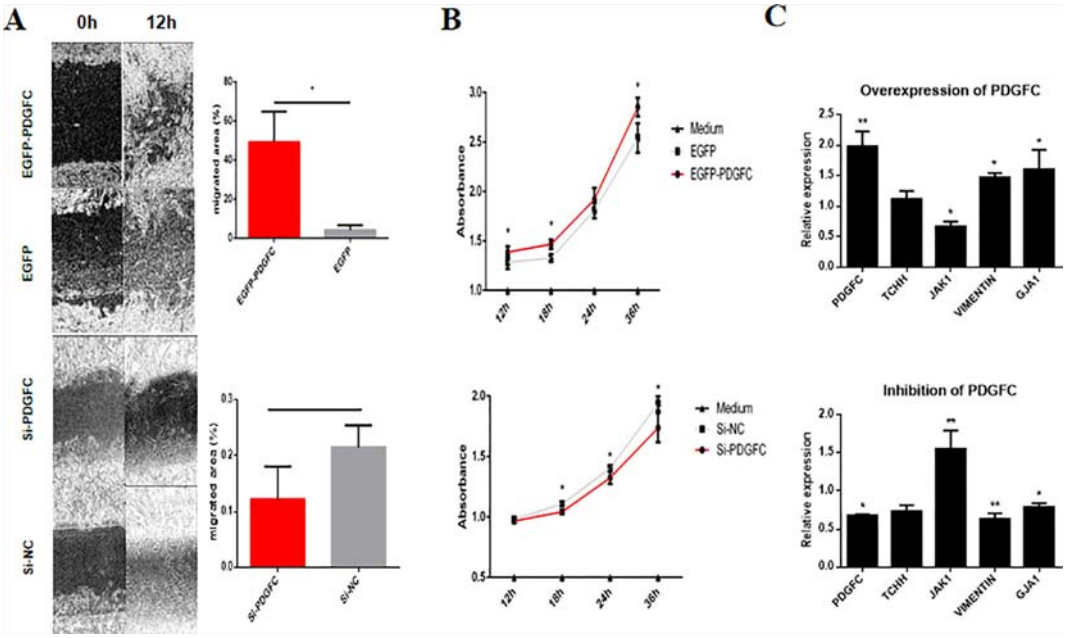


FIGURE 7 | The *PDGFC* gene affects human hair inner root sheath cell (HHIRSC) migration and proliferation by regulating hair follicle-related gene expression. **(A)** A wound healing assay was conducted after the transient transfection of *PDGFC* vectors or si-*PDGFC*. Wound closure was monitored after 12 h, and the calculation of the wound closure area is explained in the methods. The mean ± standard deviation (n = 3), $^{*}P < 0.05$ (two-sided t-test). **(B)** The cell proliferation rate was evaluated at four time points using the Cell Counting Kit-8 assay after the corresponding cell treatments. The mean ± standard deviation (n = 3), $^{*}P < 0.05$, (two-sided t-test). **(C)** Results from the quantitation of gene expression related to hair development by quantitative PCR in HHIRSCs at 48 h after transfection with control (scrambled) or the *PDGFC* vector or si-*PDGFC*. N = 3; $^{*}P < 0.05$, $^{**}P < 0.01$ (two-sided t-test).

GO terms associated with the DEGs were acquired, and the genes related to epidermal cell development were included in the five most significant terms, which intriguingly, included genes related to skin development (GO: 0043588, *TGM5, KRT23, PTCH2, S100A7, GRHL3, KRT84, MYSM1, KRT80, ACER1, KRT72, KRT2, LIPM, DSG4, SPRR4, KRTAP15-1, NF1, KRT40,* and *KRTAP3-1*) and cornification (GO: 0070268, *TGM5, KRT23, KRT84, KRT80, KRT72, KRT2, LIPM, DSG4,* and *KRT40*) (**Supplementary Table S4**), which had been closely associated with the formation of curly HFs or hair shafts in previous studies (Westgate et al., 2017). EGFR tyrosine kinase inhibitor resistance and the PI3K-Akt signaling pathway were enriched in the WGCNA modules calculated, which have associations with wavy hair coat and curly whiskers in mice (Cheng et al., 2010) and the hair cycle in many species (Kobielak et al., 2007; Feutz et al., 2008; Nie et al., 2018), respectively.

In the present study, we discovered that exonic regions showed relatively higher methylation levels compared with other regions in all nucleotide contexts (CpG, CHG, and CHH sites) (**Figure 4C**). Additionally, altered methylation profiles may induce changes to the mediation of alternative splicing through methyl-binding domain proteins (MBDs), which can regulate splicing factors indirectly (Gelfman et al., 2013). Therefore, it is reasonable to conclude that DNA methylation may regulate early HF development by mediating RNA expression and expanding the coding capacity of genes; however, this hypothesis still needs to be supported through further exploration. Among the 3,379 significant DMRs, 1,250 DMR-related genes were identified, the majority of which annotated at intronic regions (**Figure 5B**). The GO functional analysis in our study demonstrated that these DMGs were mainly enriched in the classifications of cell structure (e.g., cell projection and plasma membrane) and cell communication (e.g., cytoskeletal protein binding, channel activity and ion channel activity) (**Supplementary Figure S4**). Cell junctions (GO: 0030054) and cytoskeletal protein binding (GO: 0008092) have been reported to regulate cortex cell movement and reshaping in certain areas of the HF (Morioka et al., 2006; Harland and Plowman, 2018). These results were in agreement with previous reports and verified the importance of intercellular communication during cell reshaping of the HF bulb (Runswick et al., 2001; Arita et al., 2004).

To characterize the correlation between gene methylation and expression levels, we further focused on identifying differentially methylated (DMR-associated) and DEGs through DNA methylation profile and RNA-seq data. This integrated analysis led to the identification of 14 overlapping genes (*MFSD6, SMAD3, DICER1, THADA, ABCC11, BLMH, LGMN, NOD1, NME7, CCDC91, MAP2, ATP13A5, SIK3,* and *PDGFC*) (**Figure 6A**). Based on the KEGG pathway analysis of the DEGs and DMGs, we found that *PDGFC* and *SMAD3* were involved in the signaling pathways identified, which include roles in gap junction, EGFR tyrosine kinase inhibitor resistance, TGF-beta signaling and adherens junction, that are essential for proliferation, differentiation, and communication of HF cells during movement (Young et al., 2003; Arita et al., 2004; Plasari et al., 2010; Oshimori and Fuchs, 2012; Gay et al., 2015; Flores

et al., 2018). According to the PPI network, regulated correlation between overlapping genes was detected, with 4 of 14 overlapping genes retained in the network. Among them, SMAD family members, including *SMAD2*, *SMAD3*, *SMAD4*, and *SMAD7*, have a strong relationship to protein regulation and have been reported and validated to function in the initiation of HF cycling through conventional signaling pathways (Alimperti et al., 2012; Oshimori and Fuchs, 2012; Wang et al., 2017). The overlapping gene *NOD1*, a gene implicated in the immune response that is known to obstruct bacterial invasion and initiate inflammation, as well as exert inhibitory effects on tumor cell viability and proliferation (Velloso et al., 2018).

For validation of the downstream effects, we found that higher *PDGFC* expression promoted cell migration and proliferation, a finding that is consistent with the conclusion that tight HF inner root sheath structure surrounding the HF cortical cells can promote hair growth (Langbein et al., 2003; Basmanav FB et al., 2016). The overexpression/suppression of *PDGFC* in HHIRSC also altered the RNA expression of four key factors (*TCHH*, *JAK1*, *VIMENTIN*, and *GJA1*) that are involved in related HF-regulating pathways. The mesenchymal marker *VIMENTIN* was upregulated in *PDGFC*-overexpressing HHIRSCs, a finding that supports the result from the cell proliferation assay showing increasing proliferation rates (**Figures 7B**, **C**). As a factor in EGFR tyrosine kinase inhibitor resistance signaling, *JAK1* is a molecule known to perturb skin hemostasis-related pathways and induce epidermal inflammation and downstream genes associated with clinical skin diseases (Jin et al., 2014; Jabbari et al., 2015; Yasuda et al., 2016). *GJA1* encodes the gap junction protein connexin 43 (Cx43) in nearly every tissue in the body, and high expression leads to increased connexin density, which has been observed in the proximal bulb of the IRS (Flores et al., 2018). Intriguingly, we did not find a significant expression change in the expression level of *TCHH*, which is a key gene that interacts with IRS keratins that contributes to the "hardening" process that molds hair fiber shape. We hypothesized that the low differentiation level of the HHIRSCs in our study resulted in the lack of keratin production, thus affecting the initiation of *TCHH* expression, as *TCHH* only appears where a hardened keratin structure is needed (O'Keefe et al., 1993). These results highlight the potential role of *PDGFC* in interacting with key regulators of HF development and in initiating epidermal cell proliferation to complete HF structure.

As our DNA methylomes and transcript profiles are derived only from shoulder skin showing significant trends in fur structure, other parts of the skin tissues should be further assessed by taking measurements at more time points and by excluding possible interfering factors during the developmental stage. In addition, the effect of *PDGFC* targeted to the HF formation signatures on the hair shape transition *in vivo,* together with our data, may explain these dynamic fiber changes more convincingly.

CONCLUSIONS

This study determined the role of methylation dynamics in the curly fleece transition of the Chinese Zhongwei goat. The profile of DNA methylation and gene expression was affected among the kids during postnatal development, suggesting that epigenetic processes contribute to the developmental transitions largely driven by regulating related biological factors. We identified 1,250 DMGs that mainly function in adherens junction and gap junctions and 14 of these DMGs were differentially expressed. Importantly, among 14 overlapping genes, the *PDGFC* gene was implicated in this study as a potentially important molecule in hair formation, and the validation of the supposition in vitro demonstrates that *PDGFC* plays a significant role in regulating HHIRSC proliferation and migration. The data presented here highlights the importance of epigenetic mechanisms in molding hair shape in Zhongwei fur goats, making their fleece dynamics a promising model for the determination of hair shape and curliness in humans.

DATA AVAILABILITY STATEMENT

The datasets generated in this paper can be found at Sequence Read Archive: PRJNA524985.

ETHICS STATEMENT

All of the animal experimental procedures were performed in accordance with the guidelines for the care and use of experimental animals established by the Ministry of Agriculture of the People's Republic of China and approved by the Institutional Animal Care and Use Committee at the Institute of Animal Science, Chinese Academy of Agricultural Sciences.

AUTHOR CONTRIBUTIONS

QZ conceived and designed the experiments and revised the manuscript. PX performed the experiments, analyzed the data and wrote the manuscript. TZ and YD analyzed the data and were involved in sample collection. YM revised the manuscript. ZL, WG, XH, YP and LJ participated in the collection of samples. PX and TZ contributed equally.

FUNDING

The authors declare that this study received funding from the Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (ASTIP-IAS01) and the Modern wool sheep industry system (CARS-39-01). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01263/full#supplementary-material

REFERENCES

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacon-Duque, J. C., et al. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nat. Commun. no* 7, 10815. doi: 10.1038/ncomms10815

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13 (10), R87. doi: 10.1186/gb-2012-13-10-r87

Alimperti, S., Lei, P., Tian, J., and Andreadis, S. T. (2012). A novel lentivirus for quantitative assessment of gene knockdown in stem cell differentiation. *Gene Ther.* 19 (12), 1123–1132. doi: 10.1038/gt.2011.208

Arita, K., Akiyama, M., Tsuji, Y., McMillan, J. R., Eady, R. A., and Shimizu, H. (2004). Gap junction development in the human fetal hair follicle and bulge region. *Br. J. Dermatol.* 150 (3), 429–434. doi: 10.1046/j.1365-2133.2004.05775.x

Basmanav FB, U., Cau, L., Tafazzoli, A., Mechin, M. C., Wolf, S., Romano, M. T., et al. (2016). Mutations in three genes encoding proteins involved in hair shaft formation cause uncombable hair syndrome. *Am. J. Hum. Genet.* 99 (6), 1292–1304. doi: 10.1016/j.ajhg.2016.10.004

Bock, C., Beerman, I., Lien, W. H., Smith, Z. D., Gu, H., Boyle, P., et al. (2012). DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. *Mol. Cell* 47 (4), 633–647. doi: 10.1016/j.molcel.2012.06.019

Botchkarev, V. A., Fessing, M. Y., Botchkareva, N. V., Westgate, G., and Tobin, D. J. (2013). First International Symposium "Epigenetic Control of Skin Development and Regeneration:" how chromatin regulators orchestrate skin functions. *J. Invest. Dermatol.* 133 (8), 1918–1921. doi: 10.1038/jid.2013.126

Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., et al. (2013). InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res.* 41 (Database issue), D1228–D1233. doi: 10.1093/nar/gks1147

Cheng, X., Jin, J., Hu, L., Shen, D., Dong, X. P., Samie, M. A., et al. (2010). TRP channel regulates EGFR signaling in hair morphogenesis and skin barrier formation. *Cell* 141 (2), 331–343. doi: 10.1016/j.cell.2010.03.013

Demars, J., Cano, M., Drouilhet, L., Plisson-Petit, F., Bardou, P., Fabre, S., et al. (2017). Genome-wide identification of the mutation underlying fleece variation and discriminating ancestral hairy species from modern woolly sheep. *Mol. Biol. Evol.* 34 (7), 1722–1729. doi: 10.1093/molbev/msx114

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31 (2), 135–141. doi: 10.1038/nbt2478

Espada, J., Carrasco, E., and Calvo, M. I. (2014). Standard DNA methylation analysis in mouse epidermis: bisulfite sequencing, methylation-specific PCR, and 5-methyl-cytosine (5mC) immunological detection. *Methods Mol. Biol.* 1094, 221–231. doi: 10.1007/978-1-62703-706-8_17

Fan, Y. X., Wu, R. B., Qiao, X., Zhang, Y. J., Wang, R. J., Su, R., et al. (2015). Hair follicle transcriptome profiles during the transition from anagen to catagen in Cashmere goat (Capra hircus). *Genet. Mol. Res.* 14 (4), 17904–17915. doi: 10.4238/2015.December.22.15

Feutz, A. C., Barrandon, Y., and Monard, D. (2008). Control of thrombin signaling through PI3K is a mechanism underlying plasticity between hair follicle dermal sheath and papilla cells. *J. Cell Sci.* 121 (Pt 9), 1435–1443. doi: 10.1242/jcs.018689

Flores, A. F., Varela-Vazquez, A., Mayan, M. D., and Fonseca, E. (2018). Expression of connexin 43 in the human hair follicle: emphasis on the connexin 43 protein levels in the bulge and through the keratinization process. *J. Cutan Pathol.* 45 (1), 8–15. doi: 10.1111/cup.13050

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33 (3), 243–246. doi: 10.1038/nbt3172

Gao, Y., Wang, X., Yan, H., Zeng, J., Ma, S., Niu, Y., et al. (2016). Comparative transcriptome analysis of fetal skin reveals key genes related to hair follicle morphogenesis in cashmere goats. *PloS One* 11 (3), e0151118. doi: 10.1371/journal.pone.0151118

Gay, D. L., Yang, C. C., Plikus, M. V., Ito, M., Rivera, C., Treffeisen, E., et al. (2015). CD133 expression correlates with membrane beta-catenin and E-cadherin loss from human hair follicle placodes during morphogenesis. *J. Invest. Dermatol.* 135 (1), 45–55. doi: 10.1038/jid.2014.292

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.* 23 (5), 789–799. doi: 10.1101/gr.143503.112

Glover, J. D., Wells, K. L., Matthaus, F., Painter, K. J., Ho, W., Riddell, J., et al. (2017). Hierarchical patterning modes orchestrate hair follicle morphogenesis. *PloS Biol.* 15 (7), e2002117. doi: 10.1371/journal.pbio.2002117

Gong, W. (1994). Characteristics and classification of hair follicles and hair fibers in Zhongwei goat. *Acta Veterinaria et Zootechnica Sinica* 25 (4), 323–328.

Gudjonsson, J. E., and Krueger, G. (2012). A role for epigenetics in psoriasis: methylated Cytosine-Guanine sites differentiate lesional from nonlesional skin and from normal skin. *J. Invest. Dermatol.* 132 (3 Pt 1), 506–508. doi: 10.1038/jid.2011.364

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., et al. (2014). The DNA methylation landscape of human early embryos. *Nature* 511 (7511), 606–610. doi: 10.1038/nature13544

Harland, D. P., and Plowman, J. E. (2018). Development of hair fibres. *Adv. Exp. Med. Biol.* 1054, 109–154. doi: 10.1007/978-981-10-8195-8_10

Hynd, P. I., Edwards, N. M., Hebart, M., McDowall, M., and Clark, S. (2009). Wool fibre crimp is determined by mitotic asymmetry and position of final keratinisation and not ortho- and para-cortical cell segmentation. *Animal* 3 (6), 838–843. doi: 10.1017/S1751731109003966

Jabbari, A., Dai, Z., Xing, L., Cerise, J. E., Ramot, Y., Berkun, Y., et al. (2015). Reversal of alopecia areata following treatment with the JAK1/2 inhibitor baricitinib. *EBioMedicine* 2 (4), 351–355. doi: 10.1016/j.ebiom.2015.02.015

Jin, S. H., Choi, D., Chun, Y. J., and Noh, M. (2014). Keratinocyte-derived IL-24 plays a role in the positive feedback regulation of epidermal inflammation in response to environmental and endogenous toxic stressors. *Toxicol. Appl. Pharmacol.* 280 (2), 199–206. doi: 10.1016/j.taap.2014.08.019

Kang, X., Liu, G., Liu, Y., Xu, Q., Zhang, M., and Fang, M. (2013). Transcriptome profile at different physiological stages reveals potential mode for curly fleece in Chinese tan sheep. *PloS One* 8 (8), e71763. doi: 10.1371/journal.pone.0071763

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth3317

Kobielak, K., Stokes, N., de la Cruz, J., Polak, L., and Fuchs, E. (2007). Loss of a quiescent niche but not follicle stem cells in the absence of bone morphogenetic protein signaling. *Proc. Natl. Acad. Sci. U. S. A.* 104 (24), 10063–10068. doi: 10.1073/pnas.0703004104

Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27 (11), 1571–1572. doi: 10.1093/bioinformatics/btr167

Langbein, L., Rogers, M. A., Praetzel, S., Winter, H., and Schweizer, J. (2003). K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle. *J. Invest. Dermatol.* 120 (4), 512–522. doi: 10.1046/j.1523-1747.2003.12087.x

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Legrand, J. M. D., Roy, E., Ellis, J. J., Francois, M., Brooks, A. J., and Khosrotehrani, K. (2016). STAT5 activation in the Dermal Papilla is important for hair follicle growth phase induction. *J. Invest. Dermatol.* 136 (9), 1781–1791. doi: 10.1016/j.jid.2016.04.014

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, C., Li, Y., Zhou, G., Gao, Y., Ma, S., Chen, Y., et al. (2018). Whole-genome bisulfite sequencing of goat skins identifies signatures associated with hair cycling. *BMC Genomics* 19 (1), 638. doi: 10.1186/s12864-018-5002-5

Lu, C. P., Polak, L., Keyes, B. E., and Fuchs, E. (2016). Spatiotemporal antagonism in mesenchymal-epithelial signaling in sweat versus hair fate decision. *Science* 354 (6319), 1551. doi: 10.1126/science.aah6102

Morgenthaler, C., Diribarne, M., Capitan, A., Legendre, R., Saintilan, R., Gilles, M., et al. (2017). A missense variant in the coil1A domain of the keratin 25 gene is associated with the dominant curly hair coat trait (Crd) in horse. *Genet. Sel Evol.* 49 (1), 85. doi: 10.1186/s12711-017-0359-5

Morioka, K., Matsuzaki, T., and Takata, K. (2006). Localization of myosin and actin in the pelage and whisker hair follicles of rat. *Acta Histochem Cytochem* 39 (4), 113–123. doi: 10.1267/ahc.06004

Nie, Y., Li, S., Zheng, X., Chen, W., Li, X., Liu, Z., et al. (2018). Transcriptome reveals long non-coding RNAs and mRNAs involved in primary wool follicle induction in carpet sheep fetal skin. *Front. Physiol.* 9, 446. doi: 10.3389/fphys.2018.00446

Nishio, Y., Noguchi, E., Ito, S., Ichikawa, E., Umebayashi, Y., Otsuka, F., et al. (2001). Mutation and association analysis of the interferon regulatory factor 2 gene (IRF2) with atopic dermatitis. *J. Hum. Genet.* 46 (11), 664–667. doi: 10.1007/s100380170018

O'Keefe, E. J., Hamilton, E. H., Lee, S. C., and Steinert, P. (1993). Trichohyalin: a structural protein of hair, tongue, nail, and epidermis. *J. Invest. Dermatol.* 101 (1 Suppl), 65S–71S. doi: 10.1111/1523-1747.ep12362866

Oshimori, N., and Fuchs, E. (2012). Paracrine TGF-beta signaling counterbalances BMP-mediated repression in hair follicle stem cell activation. *Cell Stem Cell* 10 (1), 63–75. doi: 10.1016/j.stem.2011.11.005

Petridis, C., Navarini, A. A., Dand, N., Saklatvala, J., Baudry, D., Duckworth, M., et al. (2018). Genome-wide meta-analysis implicates mediators of hair follicle development and morphogenesis in risk for severe acne. *Nat. Commun.* 9 (1), 5075. doi: 10.1038/s41467-018-07459-5

Plasari, G., Edelmann, S., Hogger, F., Dusserre, Y., Mermod, N., and Calabrese, A. (2010). Nuclear factor I-C regulates TGF-{beta}-dependent hair follicle cycling. *J. Biol. Chem.* 285 (44), 34115–34125. doi: 10.1074/jbc.M110.120659

Plikus, M. V., Guerrero-Juarez, C. F., Treffeisen, E., and Gay, D. L. (2015). Epigenetic control of skin and hair regeneration after wounding. *Exp. Dermatol.* 24 (3), 167–170. doi: 10.1111/exd.12488

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., et al. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 44 (W1), W83–W89. doi: 10.1093/nar/gkw199

Rognoni, E., Widmaier, M., Jakobson, M., Ruppert, R., Ussar, S., Katsougkri, D., et al. (2014). Kindlin-1 controls Wnt and TGF-beta availability to regulate cutaneous stem cell proliferation. *Nat. Med.* 20 (4), 350–359. doi: 10.1038/nm3490

Runswick, S. K., O'Hare, M. J., Jones, L., Streuli, C. H., and Garrod, D. R. (2001). Desmosomal adhesion regulates epithelial morphogenesis and cell positioning. *Nat. Cell Biol.* 3 (9), 823–830. doi: 10.1038/ncb0901-823

Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L., and Khavari, P. A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* 463 (7280), 563–567. doi: 10.1038/nature08683

Sennett, R., Wang, Z., Rezza, A., Grisanti, L., Roitershtein, N., Sicchio, C., et al. (2015). An integrated transcriptome atlas of embryonic hair follicle progenitors, their niche, and the developing skin. *Dev. Cell* 34 (5), 577–591. doi: 10.1016/j.devcel.2015.06.023

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* 111 (51), E5593–E5601. doi: 10.1073/pnas.1419161111

Sriwiriyanont, P., Hachiya, A., Pickens, W. L., Moriwaki, S., Kitahara, T., Visscher, M. O., et al. (2011). Effects of IGF-binding protein 5 in dysregulating the shape of human hair. *J. Invest. Dermatol.* 131 (2), 320–328. doi: 10.1038/jid.2010.309

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Telerman, S. B., Rognoni, E., Sequeira, I., Pisco, A. O., Lichtenberger, B. M., Culley, O. J., et al. (2017). Dermal Blimp1 acts downstream of epidermal TGFbeta and Wnt/beta-Catenin to regulate hair follicle formation and growth. *J. Invest. Dermatol.* 137 (11), 2270–2281. doi: 10.1016/j.jid.2017.06.015

Tripurani, S. K., Wang, Y., Fan, Y. X., Rahimi, M., Wong, L., Lee, M. H., et al. (2018). Suppression of Wnt/beta-catenin signaling by EGF receptor is required for hair follicle development. *Mol. Biol. Cell* 29 (22), 2784–2799. doi: 10.1091/mbc.E18-08-0488

Velloso, F. J., Sogayar, M. C., and Correa, R. G. (2018). Expression and *in vitro* assessment of tumorigenicity for NOD1 and NOD2 receptors in breast cancer cell lines. *BMC Res. Notes* 11 (1), 222. doi: 10.1186/s13104-018-3335-4

Wang, S., Ge, W., Luo, Z., Guo, Y., Jiao, B., Qu, L., et al. (2017). Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (Capra hircus). *BMC Genomics* 18 (1), 767. doi: 10.1186/s12864-017-4145-0

Westgate, G. E., Ginger, R. S., and Green, M. R. (2017). The biology and genetics of curly hair. *Exp. Dermatol.* 26 (6), 483–490. doi: 10.1111/exd.13347

Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 34 (Web Server issue), W720–W724. doi: 10.1093/nar/gkl167

Yang, M., Song, S., Dong, K., Chen, X., Liu, X., Rouzi, M., et al. (2017). Skin transcriptome reveals the intrinsic molecular mechanisms underlying hair follicle cycling in Cashmere goats under natural and shortened photoperiod conditions. *Sci. Rep.* 7 (1), 13502. doi: 10.1038/s41598-017-13986-w

Yasuda, T., Fukada, T., Nishida, K., Nakayama, M., Matsuda, M., Miura, I., et al. (2016). Hyperactivation of JAK1 tyrosine kinase induces stepwise, progressive pruritic dermatitis. *J. Clin. Invest.* 126 (6), 2064–2076. doi: 10.1172/JCI82887

Young, P., Boussadia, O., Halfter, H., Grose, R., Berger, P., Leone, D. P., et al. (2003). E-cadherin controls adherens junctions in the epidermis and the renewal of hair follicles. *EMBO J.* 22 (21), 5723–5733. doi: 10.1093/emboj/cdg560

Yu, Z., Wildermoth, J. E., Wallace, O. A., Gordon, S. W., Maqbool, N. J., Maclean, P. H., et al. (2011). Annotation of sheep keratin intermediate filament genes and their patterns of expression. *Exp. Dermatol.* 20 (7), 582–588. doi: 10.1111/j.1600-0625.2011.01274.x

Zhao, M., Liang, G., Wu, X., Wang, S., Zhang, P., Su, Y., et al. (2012). Abnormal epigenetic modifications in peripheral blood mononuclear cells from patients with alopecia areata. *Br. J. Dermatol.* 166 (2), 226–273. doi: 10.1111/j.1365-2133.2011.10646.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xiao, Zhong, Liu, Ding, Guan, He, Pu, Jiang, Ma and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fgene.2019.01308

Dynamic Transcriptomic Analysis of Breast Muscle Development From the Embryonic to Post-hatching Periods in Chickens

Jie Liu[1,2,3†], *Qiuxia Lei*[1,2,3†], *Fuwei Li*[1,2,3], *Yan Zhou*[1,2,3], *Jinbo Gao*[1,2,3], *Wei Liu*[1,2,3], *Haixia Han*[1,2,3*] *and Dingguo Cao*[1,2,3*]

[1] Poultry Institute, Shandong Academy of Agricultural Sciences, Jinan, China, [2] Poultry Breeding Engineering Technology Center of Shandong Province, Jinan, China, [3] Shandong Provincial Key Laboratory of Poultry Diseases Diagnosis and Immunology, Jinan, China

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin,
Ireland

Reviewed by:
Yulin Jin,
Emory University, United States
Ed Smith,
Virginia Tech, United States

***Correspondence:**
Haixia Han
hanhaixia@163.com
Dingguo Cao
cdgjqs@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 09 July 2019
Accepted: 27 November 2019
Published: 10 January 2020

Citation:
Liu J, Lei Q, Li F, Zhou Y, Gao J, Liu W, Han H and Cao D (2020) Dynamic Transcriptomic Analysis of Breast Muscle Development From the Embryonic to Post-hatching Periods in Chickens. Front. Genet. 10:1308. doi: 10.3389/fgene.2019.01308

Skeletal muscle development and growth are closely associated with efficiency of poultry meat production and its quality. We performed whole transcriptome profiling based on RNA sequencing of breast muscle tissue obtained from Shouguang chickens at embryonic days (E) 12 and 17 to post-hatching days (D) 1, 14, 56, and 98. A total of 9,447 differentially expressed genes (DEGs) were filtered ($Q < 0.01$, fold change > 2). Time series expression profile clustering analysis identified five significantly different expression profiles that were divided into three clusters. DEGs from cluster I with downregulated pattern were significantly enriched in cell proliferation processes such as cell cycle, mitotic nuclear division, and DNA replication. DEGs from cluster II with upregulated pattern were significantly enriched in metabolic processes such as glycolysis/gluconeogenesis, insulin signaling pathway, calcium signaling pathway, and biosynthesis of amino acids. DEGs from cluster III, with a pattern that increased from E17 to D1 and then decreased from D1 to D14, mainly contributed to lipid metabolism. Therefore, this study may help us explain the mechanisms underlying the phenotype that myofiber hyperplasia occurs predominantly during embryogenesis and hypertrophy occurs mainly after birth at the transcriptional level. Moreover, lipid metabolism may contribute to the early muscle development and growth. These findings add to our knowledge of muscle development in chickens.

Keywords: breast muscle, chickens, development, differential gene expression, RNA sequencing

Abbreviations: ACADL, acyl-CoA dehydrogenase, long chain; ACADS, acyl-CoA dehydrogenase, C-2 to C-3 short chain; ACAT1, acetyl-CoA acetyltransferase 1; AUH, AU RNA binding methylglutaconyl-CoA hydratase; CCNA2, cyclin A2; CCNB2, cyclin B2; CDK1, cyclin-dependent kinase 1; ECHS1, enoyl-CoA hydratase, short chain 1; FBP2, fructose-bisphosphatase 2; HADHA, hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit; PGK2, phosphoglycerate kinase 2; PHKG1, phosphorylase kinase catalytic subunit gamma 1; PKLR, pyruvate kinase, liver and RBC; PPP1R3C, protein phosphatase 1 regulatory subunit 3C; TPI1, triosephosphate isomerase 1.

INTRODUCTION

In chicken production, skeletal muscle development is closely associated with the amount of meat production and its quality, ultimately affecting the economic benefits. Therefore, elucidating the molecular mechanisms underlying chicken skeletal muscle development is of vital interest. The muscle mass is determined by cell numbers and unit cell size. Hyperplasia refers to the increases in cell number or muscle fiber number that occur mainly in the embryonic period, as the number of muscle fibers is fixed by the day of hatching. However, hypertrophy refers to the increase in cell size that occurs mainly after birth (Ylihärsilä et al., 2007; Liu et al., 2017b; Ouyang et al., 2017). Therefore, there may be distinct molecular processes that occur in chicken muscle development between the embryonic and post-hatching periods.

Over the past few years, there has been much progress in exploring the molecular mechanisms underlying muscle growth and development in chickens, but most of the studies focused on the embryonic or post-hatching period. Davis et al. (2015) characterized the transcriptome of Ross 708 chicken breast muscle at specified time points from 6 to 21 days after hatching. Li et al. (2019) explored the messenger RNA (mRNA) and microRNA (miRNA) profiles of Gushi chicken muscle tissues in the late postnatal stage (6, 14, 22, and 30 weeks). Our previous study examined the protein expression profiles in the breast muscle of Beijing-You chickens at ages 1, 56, 98, and 140 days using isobaric tags for relative and absolute quantification (Liu et al., 2016). Li et al. (2017) and Ouyang et al. (2017) explored the transcriptome and protein expression profiles in leg muscle tissues of Xinghua chicken at embryonic days (E) 11 and E16 and post-hatching day (D) 1, respectively. However, few studies have paid attention to the whole muscle development from embryonic to post-hatching periods in chickens. Only Liu et al. (2017b) investigated the proteomes of breast muscle in Cobb and Beijing-You chickens at E12, E17, D1, and D14.

Shouguang chickens that have been breed in China for 2,000 years are a dual-purpose breed with large bodies (Gao et al., 2008), which may be excellent material for studying muscle development. Therefore, we chose the critical breast muscle developmental stages in the embryonic to post-hatching periods (E12 and E17 and D1, D14, D56, and D98) of Shouguang chickens for quantitative analysis of the gene expression profile of breast muscle, which may help us explore the development-related genes expression signatures in breast muscle and its distinction between embryonic and post-hatching periods.

METHODS

Animals

Shouguang chicken eggs were obtained from the experimental farm of the Poultry Institute (PS), Shandong Academy of Agricultural Sciences (SAAS, Jinan, China). All eggs were incubated with the normal procedure and chicks were reared in cages using standard conditions of temperature, humidity, and ventilation at the farm of the PS, SAAS. The same diet was fed to all chickens and a three-phase feeding system was used: starter ration (days 1–28) with 21.0% crude protein and 12.12 MJ/kg; second phase (days 28–56) with 19.0% crude protein and 12.54 MJ/kg; and final phase (after day 56) with 16.0% crude protein and 12.96 MJ/kg. Feed and water were provided *ad libitum* during the experiment. Breast muscles were used at E12, E17, D1, D14, D56, and D98. All fresh breast muscle tissue samples were collected, frozen in liquid nitrogen, and stored at −80°C until RNA extraction. The sex of chicken embryos was identified by polymerase chain reaction (PCR) of the *CHD1* gene (Fridolfsson and Ellegren, 1999). Chickens with two bands of 600 and 450 bp were born as female, and those with one band of 600 bp were born as male.

RNA Extraction, cRNA Library Construction, and Sequencing

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Three female chickens at each stage (except E17) were used for further experiments. Total RNA quantity and purity were analyzed using a Bioanalyer 2100 (Agilent, Santa Clara, CA, USA) with RNA integrity number >7.0. Approximately 10 μg total RNA was used to deplete rRNA using the Epicentre Ribo-Zero Gold Kit (Illumina, San Diego, CA, USA). Following purification, the poly(A)− or poly(A)+ RNA fraction was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were reverse-transcribed to create the final complementary DNA (cDNA) library in accordance with the protocol for the RNA sequencing (RNA-Seq) sample preparation kit (Illumina). The average insect size for the paired-end libraries was 300 ± 50 bp. We performed paired-end sequencing on an Illumina Hiseq 4000 at LC-Bio, China.

RNA-Seq Reads Mapping and DEG Analysis

We aligned reads to the genome of Gallus_gallus 5.0 (GCA_000002315.3) using HISAT package (Kim et al., 2015), which initially removed reads based on quality information accompanying each read and then mapped the reads to the reference genome. The mapped reads of each sample were assembled using StringTie (Pertea et al., 2015). All transcriptomes from samples were merged to reconstruct a comprehensive transcriptome using perl scripts. After the final transcriptome was generated, StringTie and edgeR (Robinson et al., 2010) were used to estimate the expression levels of all transcripts. StringTie was used to perform expression level for mRNAs by calculating fragments per kilobase of transcript per million fragments mapped (FPKM). Differentially expressed genes (DEGs) were selected with log2 (fold change) > 1 or log2 (fold change) less than −1 with statistical significance ($Q < 0.01$) by R package. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy

of Sciences, and is publicly accessible at http://bigd.big.ac.cn/gsa (accession no. CRA001773).

Time Series Expression Profile Clustering

The non-parametric clustering algorithm of STEM (Short Time-Series Expression Miner, version 1.3.11) (Ernst and Bar-Joseph, 2006) was used to cluster and visualize the expression patterns of DEGs. Expression profiles of DEGs were clustered based on their log2 (FPKM values) and their correlation coefficients. The maximum unit change in model profiles between time points was adjusted to 2 and the maximum number of model profiles to 50. The statistical significance of the number of DEGs to each profile versus the expected number was computed using the algorithm proposed by Ernst and Bar-Joseph (2006).

Functional Annotation

Functional analysis of DEGs was performed using the DAVID (Database for Annotation, Visualization and Integrated Discovery) tool (http://david.abcc.ncifcrf.gov/) (Dennis et al., 2003). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using KOBAS version 3.0 (Xie et al., 2011). Gene Ontology (GO) terms and KEGG pathways with $P < 0.05$ were considered significantly enriched groups of genes possibly contributing to muscle development.

qRT-PCR Confirmation

To confirm our differential expression results, we conducted quantitative reverse transcription PCR (qRT-PCR) for six selected genes (*MYOG*, *MYH11*, *TNNI2*, *TNNT3*, *TNNC2*, and *TPM2*). The total RNA was used for first-strand cDNA synthesis using a commercial kit (TaKaRa, Dalian, China). cDNA was subsequently used for qRT-PCR analyses with an ABI 7500 Detection System (Applied Biosystems, Foster City, CA, USA) and primers designed using Primer Premier version 5.0 (PREMIER Biosoft, Palo Alto, CA, USA), as listed in **Table S1**. mRNA abundance of candidate genes was determined using the KAPA SYBR® FAST qPCR Master Mix (2×) Universal Cocktail (KAPA Biosystems, Boston, MA, USA). qRT-PCR was performed following the instructions of ABI 7500 with default parameters. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) was used to calculate the relative mRNA abundance. The beta actin gene (*ACTB*) was used as the housekeeping gene. Three independent replications were used for each assay and data were presented as means ± SD.

RESULTS

Overall Assessment for Sequencing Data Mapping Statistics

To identify mRNA expressed in breast muscle tissue development of chickens, we constructed 17 cDNA libraries (E12_1, E12_2, E12_3, E17_1, E17_2, D1_1, D1_2, D1_3, D14_1, D14_2, D14_3, D56_1, D56_2, D56_3, D98_1, D98_2, and D98_3) from breast muscle samples at six developmental stages. As shown in **Table 1**, 68,431,306–103,358,850 raw reads were generated in the 17 libraries, and 65,384,366–100,882,250 clean reads were obtained after discarding adaptor sequences and low-quality reads. We mapped clean reads to chicken reference genome Gallus_gallus 5.0 and found that 84.41–90.10% of the clean reads in the libraries were mapped to the chicken reference genome (**Table 1**).

Differential Expression Analysis of Genes

In pairwise comparisons between the libraries of breast muscle at the six developmental stages, a total of 9,447 genes were differentially expressed ($Q < 0.01$, fold change > 2) (**Figure 1** and **Table S2**). There were 2,502, 4,582, 4,394, 3,689, and 4,607 DEGs in E17, D1, D14, D56, and D98 compared to E12. Comparing successive ages within each region, 2,502, 2,429, 1,839, 262, and 144 DEGs were found in E17 versus E12, D1 versus E17, D14 versus D1, D56 versus D14, and D98 versus D56, respectively. The numbers of DEGs were greatest in E17 versus E12 and lowest in D98 versus D56, which indicated that regional differences in gene expression were greatest during the earlier stages of embryo development.

TABLE 1 | Overview of raw data output and quality assessment.

Sample	Raw reads	Clean reads	Mapped reads	Mapped rate (%)
E12_1	75,233,696	72,137,830	63,804,708	88.45
E12_2	71,868,594	68,692,864	60,781,832	88.48
E12_3	73,728,758	70,509,198	61,614,305	87.38
E17_1	76,625,816	73,423,116	65,037,045	88.58
E17_3	68,431,306	65,384,366	56,147,755	85.87
D1_1	75,819,908	71,918,030	61,870,757	86.03
D1_2	87,249,520	80,288,708	71,793,909	89.42
D1_3	75,927,124	71,486,238	60,974,309	85.30
D14_1	88,752,924	86,540,700	77,973,761	90.10
D14_2	81,110,076	79,280,896	69,851,415	88.11
D14_3	88,685,236	86,822,844	74,453,649	85.75
D56_1	86,677,672	84,671,272	74,896,179	88.46
D56_2	88,218,208	86,364,222	75,789,517	87.76
D56_3	96,7881,00	94,729,644	82,666,195	87.27
D98_1	96,819,270	94,753,276	83,684,517	88.32
D98_2	103,358,850	100,882,250	86,849,158	86.09
D98_3	89,320,116	87,415,674	73,785,643	84.41

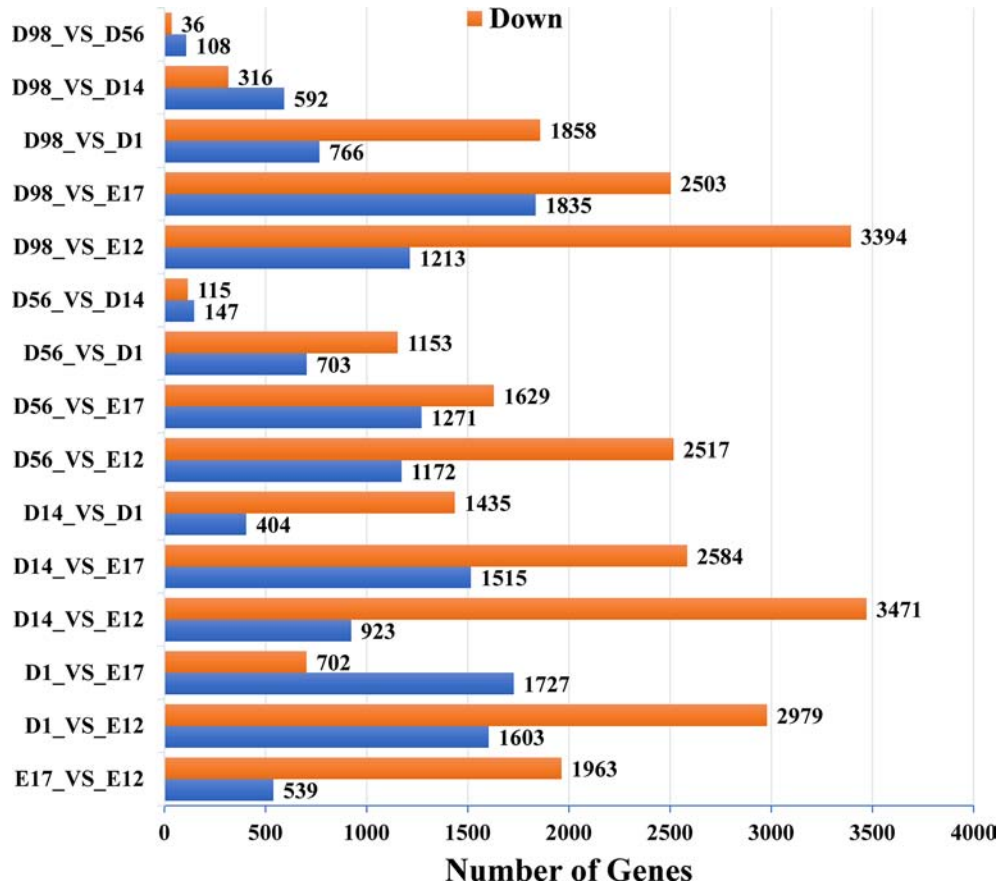


FIGURE 1 | Numbers of upregulated and downregulated genes in chicken breast muscle through pairwise comparisons.

STEM Analysis of DEG Profiles

As our data were collected at different time points, STEM was used to cluster and visualize possible changes in the profiles of 9,447 DEGs at six time points of breast muscle development. Within the 50 model profiles, five expression profiles containing 5,269 genes were statistically significant ($P < 0.05$; **Figure 2A** and **Table S3**). Of these, profiles 8 and 12 with downregulated patterns contained 3,233 and 693 DEGs, respectively (**Figure 2B** and **Table S3**), while profiles 39 and 49 with upregulated patterns contained 380 and 156 DEGs, respectively (**Figure 2C** and **Table S3**). Profile 25 with 717 genes as the third pattern showed an increase from E17 to D1 and reached a peak at D1, then decreased from D1 to D14 and remained stable from D14 to D98 (**Figure 2D** and **Table S3**). Thus, the expression pattern of DEGs can be divided into three clusters: cluster I (profiles 8 and 12, total of 3,926 DEGs) with downregulated pattern; cluster II (profiles 39 and 49, total of 536 DEGs) with upregulated pattern; and cluster III (profile 25, total of 717 DEGs). The results provide new information related to further characterization of novel molecules associated with skeletal muscle development in chickens.

GO Enrichment Analysis

To explore the biological function of DEGs, GO enrichment analysis was performed based on cluster analysis. The genes in cluster I (profiles 8 and 12) were significantly enriched in 138 GO terms (68 under biological process, 38 under cellular component, and 32 under molecular function) (**Table S4**). Within the biological process category, the most abundant GO terms consisted of DNA replication, cell division, ATP-dependent chromatin remodeling, mitotic nuclear division, and DNA repair (**Figure 3A**). The genes in cluster II (profiles 39 and 49) were significantly enriched in 34 GO terms (16 under biological process, six under cellular component, and 12 under molecular function) (**Table S5**). Within the biological process category, the most abundant GO terms consisted of carbohydrate metabolic process, xanthine catabolic process, gluconeogenesis, and positive regulation of interferon-γ production (**Figure 3B**). Skeletal muscle contraction was also included in this category (**Figure 3B**). We also analyzed the biological function of genes in cluster III (profile 25). Eighteen GO terms (10 under biological process, three under cellular component, and five under molecular function) were significantly enriched (**Table S6**).

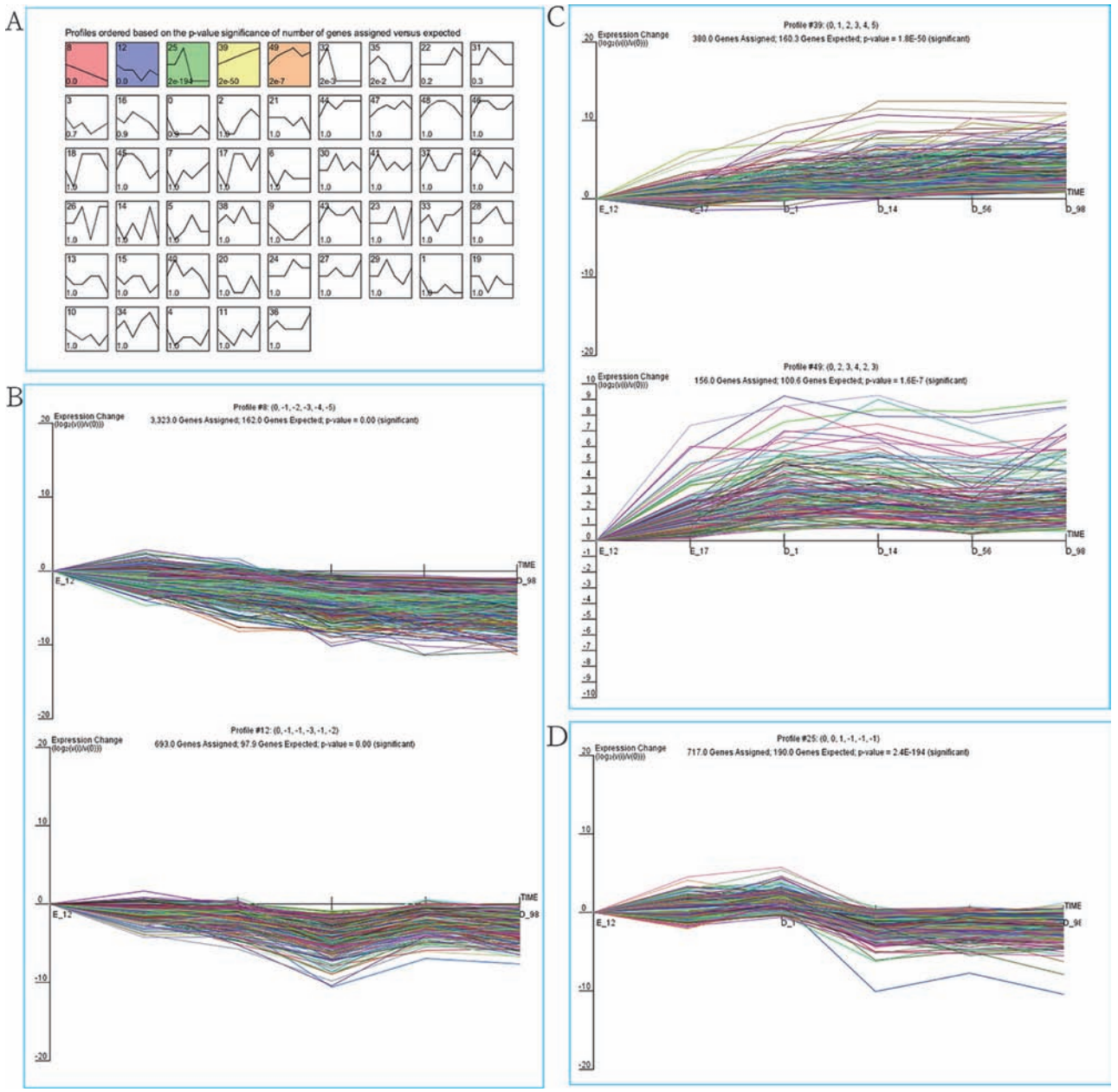


FIGURE 2 | STEM analysis of DEG profiles. **(A)** Each *box* corresponds to a type expression profile and only *colored profiles* are significantly different. The *upper-left* and *upper-right numbers in each box* indicate the order of profiles and *P* values, respectively. **(B)** Profile 8 (up) and profile 12 (down) with downregulated patterns. **(C)** Profile 39 (up) and profile 49 (down) with upregulated patterns. **(D)** Profile 25.

Within the biological process category, the most abundant GO terms consisted of fatty acid beta-oxidation using acyl-CoA dehydrogenase, fatty acid beta-oxidation, positive regulation of focal adhesion assembly, lipid homeostasis, and regulation of stress fiber assembly (**Figure 3C**).

KEGG Enrichment Analysis

We used KEGG pathway analysis to explore the signaling pathways of DEGs involved in cluster I (profiles 8 and 12), cluster II (profiles 39 and 49), and cluster III (profile 25). For cluster I, the DEGs were significantly enriched in 13 pathways (**Figure 4A** and **Table S7**), and half of these pathways are involved in cell division, such as cell cycle, DNA replication, nucleotide excision repair, mismatch repair, oocyte meiosis, and spliceosome. As for cluster II, the DEGs were significantly enriched in 16 pathways (**Figure 4B** and **Table S8**), and it was interesting that all of these pathways were directly or indirectly involved in the main metabolic processes of the organism, such

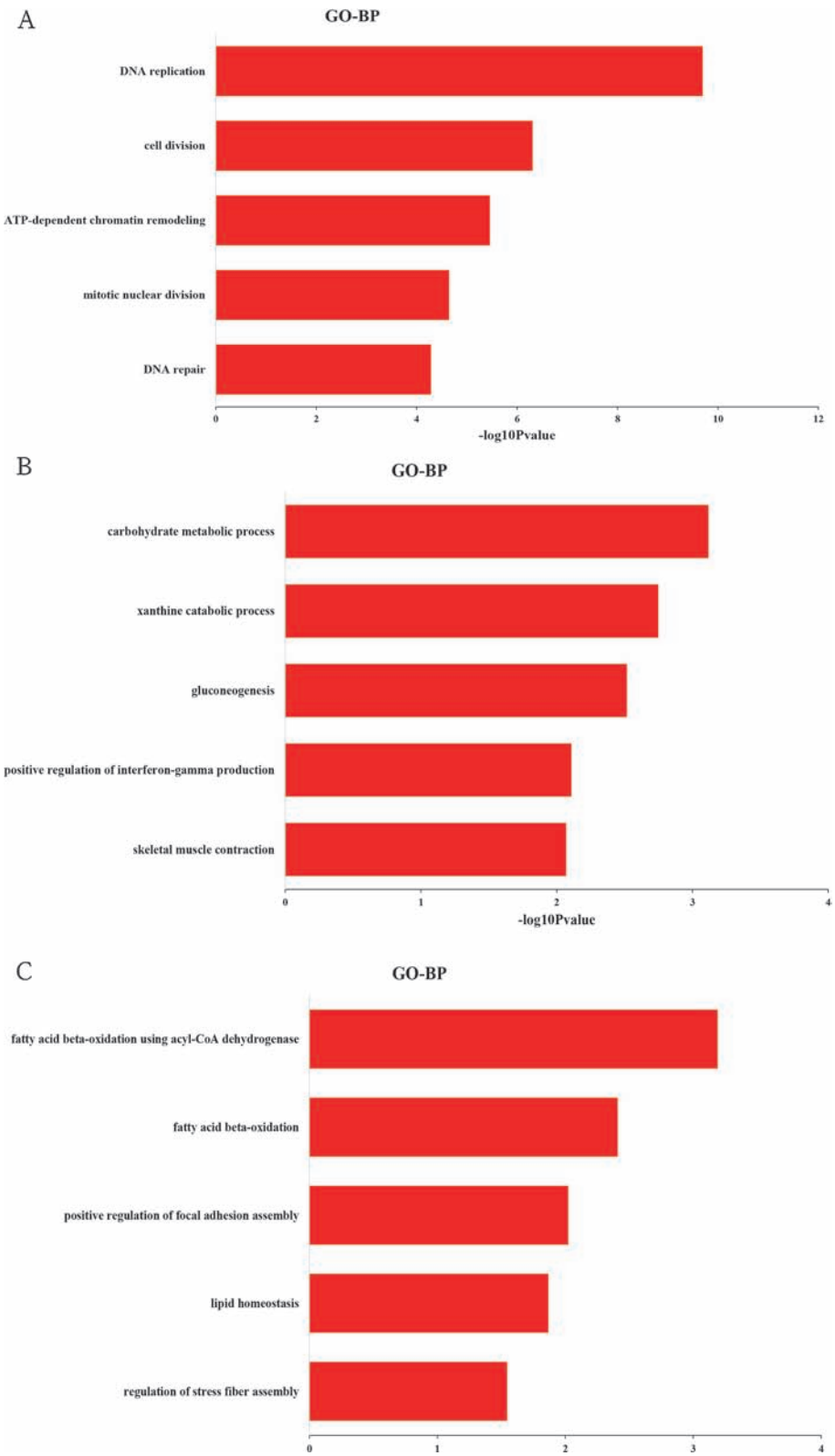


FIGURE 3 | Top 5 significantly enriched biological process terms of genes in cluster I **(A)**, cluster II **(B)**, and cluster III **(C)**.

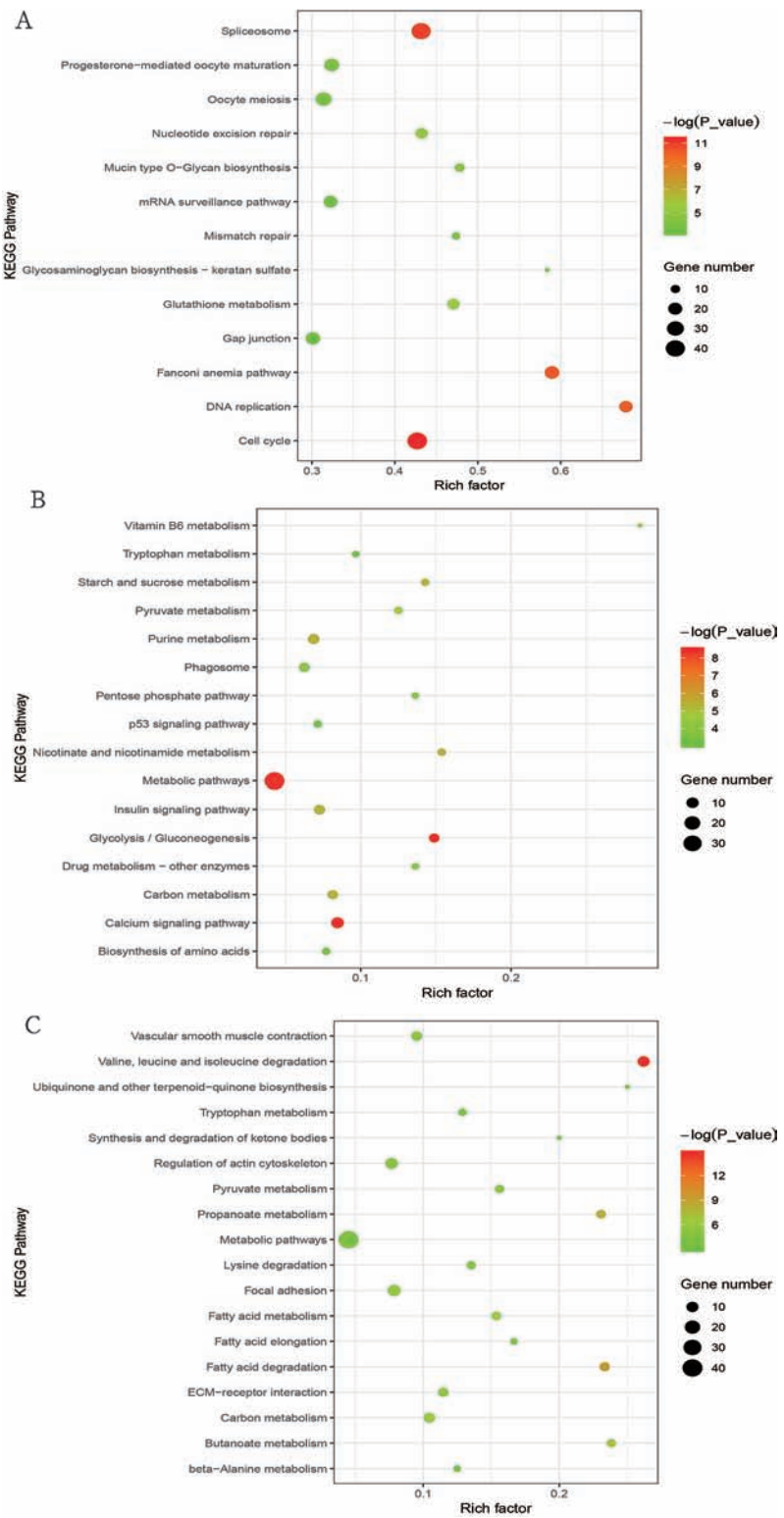


FIGURE 4 | Bubble plot of significantly enriched pathways for cluster I (profiles 8 and 12) with downregulated pattern **(A)**, cluster II (profiles 39 and 49) with upregulated pattern **(B)**, and cluster III (profile 25) **(C)**. *Bubble color* and *size* correspond to the *P* value and gene number enriched in the pathway. The rich factor indicates the ratio of the number of DEGs mapped to a certain pathway to the total number of genes mapped to this pathway.

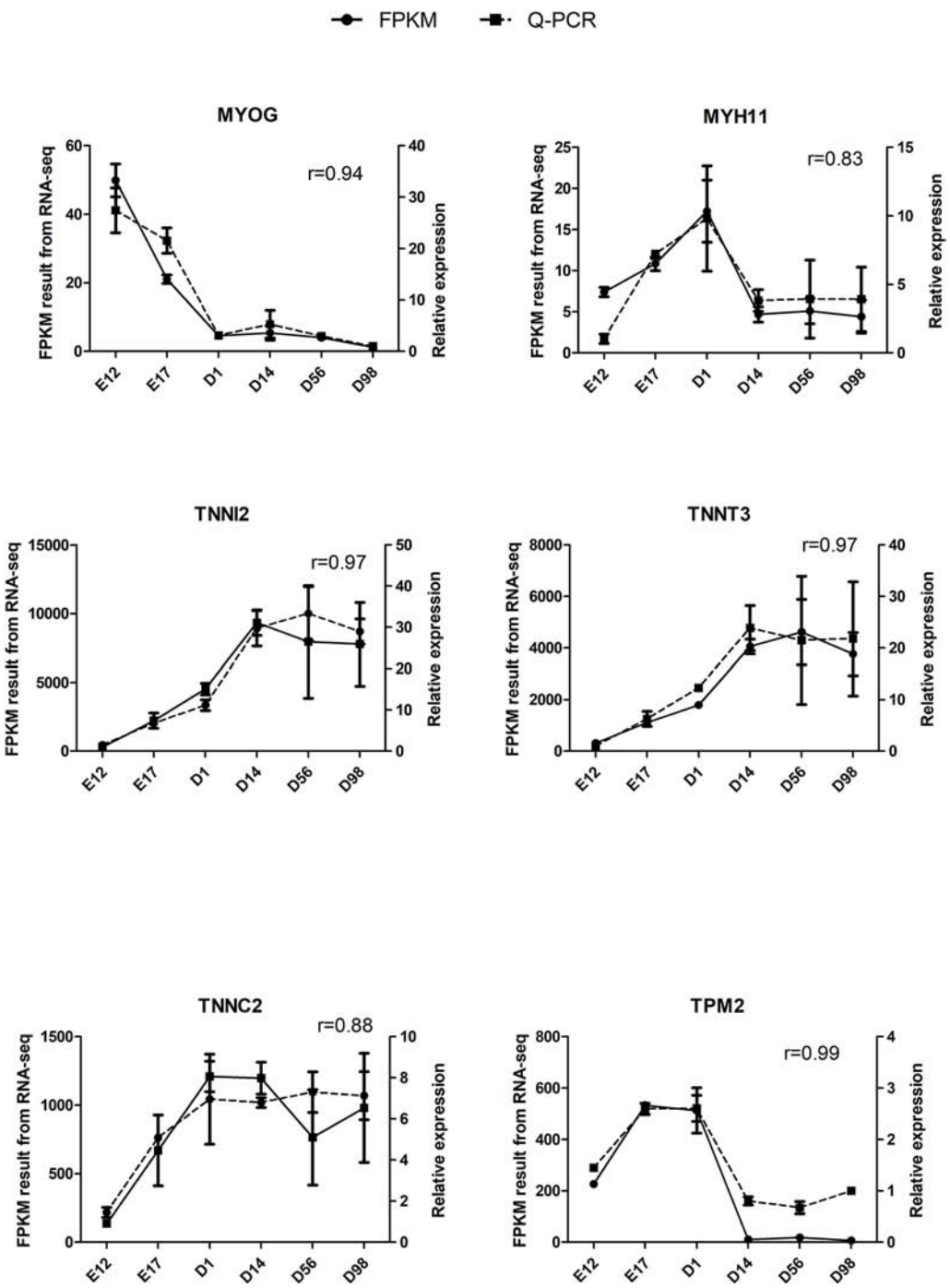


FIGURE 5 | Validation of six DEGs by qRT-PCR. The *r* value represents Pearson's correlation coefficient between two methods.

as glycolysis/gluconeogenesis, purine metabolism, starch and sucrose metabolism, vitamin B6 metabolism metabolic pathways, pentose phosphate pathway, biosynthesis of amino acids nicotinate, and insulin signaling pathway. In cluster III, the DEGs were significantly enriched in 18 pathways (**Figure 4C** and **Table S9**), and most of these pathways were associated with metabolism. Among these metabolic pathways, the most enriched pathways were those related to lipid metabolism, such

as fatty acid degradation, propanoate metabolism, butanoate metabolism, fatty acid elongation, synthesis and degradation of ketone bodies, and fatty acid metabolism.

Validation of DEGs by qRT-PCR

The qRT-PCR assays were conducted to validate six selected DEGs from RNA-Seq: *MYOG*, *MYH11*, *TNNI2*, *TNNT3*, *TNNC2*, and *TPM2*. Relative expression changes of qRT-PCR data were highly (r = 0.83–0.99) correlated with RNA-Seq data (**Figure 5**), suggesting the reliability of the RNA-Seq approach.

DISCUSSION

Skeletal muscle growth and development includes a series of closely regulated changes in gene expression level from embryo to adult, and uncovering the gene expression patterns underneath chicken skeletal muscle development contributes to meat production. Previous transcriptome analysis of chicken muscle only concentrated on the embryonic period or adult stage, and few studies have systematically examined the transcriptome of chicken skeletal muscle development from the embryonic period to the growing period. To investigate the mechanisms of skeletal muscle development systematically, we used RNA-Seq to generate extensive cDNA libraries for six developmental stages of chickens from E12 to D98. As shown in **Figure 1** and **Table S2**, a total of 9,447 DEGs were identified in pairwise comparisons between the libraries of breast muscle at the six developmental stages, and the regional differences in gene expression were greatest during the earlier stages of embryo development (E17 versus E12) than the late postnatal stage (D98 versus D56). Previous studies have demonstrated the increase in cell number or muscle fiber number which occurs mainly in embryonic periods as the numbers of muscle fibers were fixed by the day of hatching and then impacted on the postnatal accretion of muscle mass (Smith, 1963; Sporer et al., 2011). Thus, embryonic periods are the critical periods for muscle development and more genes are active in these periods. Six selected DEGs involved in muscle development were validated using qRT-PCR, and the results were consistent with those from RNA-Seq, suggesting reliability of the identified DEGs through RNA-Seq (**Figure 5**).

Since muscle development was accompanied by the differential expression of related genes in different growth periods and our data were also collected at different time points, we used STEM software, which is widely used to study dynamic biological processes (Guo et al., 2016; Ma et al., 2018; Zhan et al., 2018), to investigate the dynamic genetic changes during breast muscle development. We experimented with various numbers of profiles and found that five profiles (combining with three clusters) best captured the expression patterns of DEGs (**Figure 2** and **Table S3**). The genes in a cluster have similar temporal expression patterns and may be involved in the same biological process. Therefore, we performed GO and KEGG analyses to explore the function of the DEGs with similar temporal expression patterns.

For cluster I with downregulated pattern (**Figure 2B** and **Table S3**), 3,926 genes were significantly clustered, which was more than 40% of the total DEGs. These genes were more highly expressed in the early periods of muscle development than the late stages of growth, which further confirmed that the early development might play key roles in muscle growth. GO functional annotation and KEGG analysis both showed that the downregulated genes were significantly enriched in cell proliferation, including DNA replication, cell cycle, cell division, mitotic nuclear division, and DNA repair (**Figures 3A** and **4A**), which was similar to previous research on goat muscle development from gestation to birth which showed that genes with downregulated patterns were also involved in cell proliferation processes (Zhan et al., 2018). Therefore, these results further support the hypothesis that the total number of skeletal myofibers is defined by hyperplasia during embryogenesis. Among these downregulated genes, *CCNA2*, *CCNB2*, and *CDK1*, which encode the cyclins and their cognate cyclin-dependent protein kinases, were not only significantly enriched in the biological process of cell division but also significantly enriched in cell cycle pathway (**Tables S4** and **S7**). Cyclin A2 possesses a unique role in its two-point control of the cell cycle, first by interacting with CDK2 in controlling the G1/S transition into DNA synthesis and then by interacting with CDKs 1 and 2 to control the G2/M entry into mitosis (Li et al., 1998). Previous study has been demonstrated that constitutive expression of cyclin A2 in a transgenic mouse yields robust postnatal cardiomyocyte mitosis and hyperplasia (Chaudhry et al., 2004). Cyclin B2 was also demonstrated to have a regulatory role in chicken breast muscle development (Li et al., 2019). Moreover, CDK1 and CDK2 play integral roles in reducing MyoD activity during myoblast proliferation by phosphorylating MyoD (Kitzmann et al., 1999). These results suggest that the genes with a downregulated pattern of expression play regulatory roles in chicken breast muscle development through the processes involved in the early stages of cell proliferation, and genes related to cyclins and their cognate cyclin-dependent protein kinases may be critical factors in regulating cell proliferation.

For cluster II with an upregulated pattern of expression (**Figure 2C** and **Table S3**), the functional annotation and pathway analysis both showed that these genes were significantly enriched in metabolism such as carbohydrate metabolism, glycolysis/gluconeogenesis, calcium signaling pathway, insulin signaling pathway, and biosynthesis of amino acids (**Figures 3B** and **4B**). A previous study of goat muscle development also found that genes with upregulated patterns of expression were related to metabolic pathways, such as biosynthesis of amino acids, glycolysis/gluconeogenesis, and the TCA cycle (Zhan et al., 2018). Among these genes, *PHKG1*, *PPP1R3C*, and *FBP2* were significantly enriched not only in glycogen biosynthetic process and gluconeogenesis but also in insulin signaling pathway and calcium signaling pathway (**Tables S5** and **S8**). *PHKG1*, as a key factor in insulin signaling

and calcium signaling pathways, encodes the catalytic subunit of phosphorylase kinase, which functions in the cascade activation of glycogen breakdown in muscle tissue (Ma et al., 2014). Fructose-1,6-bisphosphatase encoded by *FBP2* catalyzes the hydrolysis of fructose-1,6-bisphosphate to fructose-6-phosphate and inorganic phosphate, which plays a regulatory role in the synthesis of glycogen/glucose. Previous findings point to FBP2 as an important link between calcium-induced muscle contractive and metabolic (glycolytic) activity, mitochondrial function, and cell survival (Pirog et al., 2014). PPP1R3C was also called protein targeting to glycogen (PTG) and regulated glycogen metabolism (Ji et al., 2019). Moreover, glycolytic process and biosynthesis of amino acids pathway were both significantly enriched in *PKLR*, *PGK2*, and *TPI1* (**Tables S5** and **S8**); thus, these genes may be important regulatory switches for protein and energy conversion and ultimately influence muscle development. A number of studies have demonstrated that muscle mass increased by hypertrophy (increased cellular protein content) after hatching and was controlled by synthesis of muscle proteins or their degradation (Braun and Gautel, 2011). Protein and energy metabolism are tightly coupled, and the energy from glycolysis/gluconeogenesis is needed for protein turnover during skeletal muscle development (Duan et al., 2016; Liu et al., 2016). Therefore, these results show that genes involved in metabolism may be critical for postnatal myofiber growth, muscle hypertrophy, and muscle regeneration, and the protein synthesis and energy metabolism of skeletal muscle regulated by insulin signaling pathway and calcium signaling pathway may be important for coordinating muscle development.

Cluster III (profile 25) showed an increase from E17 to D1 and reached a peak at D1, then decreased from D1 to D14 and remained stable from D14 to D98 (**Figure 2D** and **Table S3**). The functional annotation showed that processes and pathways involved in lipid metabolism were significantly enriched, such as fatty acid β-oxidation, lipid homeostasis, fatty acid degradation, and propanoate metabolism (**Figures 3C** and **4C**). Previous studies have demonstrated that lipids stored in the adipocytes during embryonic life are transferred to the muscle fibers and used for growth and energy requirements at the early stage, while muscle again stores lipids in later life (Chartrin et al., 2007; Liu et al., 2016). Interestingly, *ACADL*, *ACAT1*, *HADHA*, *ACADS*, *ECHS1*, and *AUH*, which are significantly enriched in fatty acid beta-oxidation, were active during embryonic life in the present study (**Table S6**), which further demonstrated that the lipids were the important energy source for muscle development and growth at the early stage. Moreover, some of these genes were also the key regulatory molecules for intramuscular fat (IMF) deposition. For example, *ACADL* and *HADHA* have been identified as candidate biomarkers for IMF deposition in Cobb and Beijing-You chickens (Liu et al., 2017a), and it was interesting that their expression patterns in Shouguang chickens were similar to Cobb and Beijing-You chickens. Moreover, *ACAT1* expression was significantly lower in muscle of AA chickens with low IMF content than in Beijing-You chickens with abundant IMF (Liu et al., 2018), suggesting that *ACAT1* may contribute to IMF deposition. These results suggest that genes involved in lipid metabolism, and especially those related to fatty acid beta-oxidation, play important roles in early muscle development and deposition of IMF.

CONCLUSION

In the present study, we systematically identified DEGs and investigated their temporal expression profiles during chicken breast muscle development from E12 to D98. A total of 9,447 DEGs were identified in chicken breast muscle and showed three significantly different expression patterns. Functional enrichment analysis suggests that genes with downregulated patterns contribute to cell proliferation processes, while genes with upregulated patterns are mainly involved in metabolism. Genes related to lipid metabolism change dramatically around the time of birth, which may play important roles in early muscle development and deposition of IMF. In summary, our study will facilitate understanding of the mechanisms underlying the phenotype that myofiber hyperplasia occurs predominantly during embryogenesis and hypertrophy occurs mainly after birth at the transcriptional level. These findings elucidate the regulatory mechanisms involved in chicken breast muscle development.

DATA AVAILABILITY STATEMENT

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences and is publicly accessible at http://bigd.big.ac.cn/gsa (accession no CRA001773).

ETHICS STATEMENT

All experiments were approved by the Animal Care Committee of the Shandong Academy of Agricultural Sciences (Ji'nan, China). The experimental procedures with chickens were performed according to the Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China).

AUTHOR CONTRIBUTIONS

JL and QL performed experiments and data analysis and draft writing. FL, YZ, JG, and WL contributed to animal experiments and data analysis. HH and DC designed experiments and supervised and coordinated the study. All authors reviewed the manuscript.

FUNDING

The research was supported by grants from Natural Science Foundation of Shandong province (ZR2019BC077), the Earmarked Fund for Modern Agro-industry Technology Research System (CARS-41), Jinan Layer Experiment Station of China Agriculture Research System (CARA-40-S12), Agricultural Scientific and Technological Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2016A04), Shandong Provincial Key Laboratory of Special Construction Project (SDKL201810), and Construction of Subjects and Teams of Institute of Poultry Science (CXGC2018E11).

ACKNOWLEDGMENTS

We thank International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01308/full#supplementary-material

REFERENCES

Braun, T., and Gautel, M. (2011). Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nat. Rev. Mol. Cell Biol.* 12 (6), 349. doi: 10.1038/nrm3118

Chartrin, P., Bernadet, M., Guy, G., Mourot, J., Hocquette, J.-F., Rideau, N., et al. (2007). Do age and feeding levels have comparable effects on fat deposition in breast muscle of mule ducks? *Animal* 1 (1), 113–123. doi: 10.1017/S1751731107658029

Chaudhry, H. W., Dashoush, N. H., Tang, H., Zhang, L., Wang, X., Wu, E. X., et al. (2004). Cyclin A2 mediates cardiomyocyte mitosis in the postmitotic myocardium. *J. Biol. Chem.* 279 (34), 35858–35866. doi: 10.1074/jbc.M404975200

Davis, R. V., Lamont, S. J., Rothschild, M. F., Persia, M. E., Ashwell, C. M., and Schmidt, C. J. (2015). Transcriptome analysis of post-hatch breast muscle in legacy and modern broiler chickens reveals enrichment of several regulators of myogenic growth. *PloS One* 10 (3), e0122525. doi: 10.1371/journal.pone.0122525

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (9), R60. doi: 10.1186/gb-2003-4-9-r60

Duan, Y., Li, F., Li, Y., Tang, Y., Kong, X., Feng, Z., et al. (2016). The role of leucine and its metabolites in protein and energy metabolism. *Amino Acids* 48 (1), 41–51. doi: 10.1007/s00726-015-2067-1

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinf.* 7 (1), 191. doi: 10.1186/1471-2105-7-191

Fridolfsson, A.-K., and Ellegren, H. (1999). A simple and universal method for molecular sexing of non-ratite birds. *J. Avian Biol.* 30 (1), 116–121. doi: 10.2307/3677252

Gao, Y., Tu, Y., Tong, H., Wang, K., Tang, X., and Chen, K. (2008). Genetic variation of indigenous chicken breeds in China and a Recessive White breed using AFLP fingerprinting. *S. Afr. J. Anim. Sci.* 38 (3), 193–200. doi: 10.4314/sajas.v38i3.4129

Guo, J., Zhao, W., Zhan, S., Li, L., Zhong, T., Wang, L., et al. (2016). Identification and expression profiling of miRNAome in goat longissimus dorsi muscle from prenatal stages to a neonatal stage. *PloS One* 11 (10), e0165764. doi: 10.1371/journal.pone.0165764

Ji, X., Wang, S., Tang, H., Zhang, Y., Zhou, F., Zhang, L., et al. (2019). PPP1R3C mediates metformin-inhibited hepatic gluconeogenesis. *Metabolism* doi: 10.1016/j.metabol.2019.06.002

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357. doi: 10.1038/nmeth.3317

Kitzmann, M., Vandromme, M., Schaeffer, V., Carnac, G., Labbé, J.-C., Lamb, N., et al. (1999). cdk1-and cdk2-mediated phosphorylation of MyoD Ser200 in growing C2 myoblasts: role in modulating MyoD half-life and myogenic activity. *Mol. Cell. Biol.* 19 (4), 3167–3176. doi: 10.1128/MCB.19.4.3167

Li, J.-M., Poolman, R. A., and Brooks, G. (1998). Role of G1 phase cyclins and cyclin-dependent kinases during cardiomyocyte hypertrophic growth in rats. *Am. J. Physiol.-Heart Circ. Physiol.* 275 (3), H814–H822. doi: 10.1152/ajpheart.1998.275.3.H814

Li, Z., Ouyang, H., Zheng, M., Cai, B., Han, P., Abdalla, B. A., et al. (2017). Integrated analysis of long non-coding RNAs (LncRNAs) and mRNA expression profiles reveals the potential role of LncRNAs in skeletal muscle development of the chicken. *Front. Physiol.* 7, 687. doi: 10.3389/fphys.2016.00687

Li, Y., Chen, Y., Jin, W., Fu, S., Li, D., Zhang, Y., et al. (2019). Analyses of MicroRNA and mRNA Expression Profiles Reveal the Crucial Interaction Networks and Pathways for Regulation of Chicken Breast Muscle Development. *Front. Genet.* 10, 197. doi: 10.3389/fgene.2019.00197

Liu, J., Fu, R., Liu, R., Zhao, G., Zheng, M., Cui, H., et al. (2016). Protein profiles for muscle development and intramuscular fat accumulation at different post-hatching ages in chickens. *PloS One* 11 (8), e0159722. doi: 10.1371/journal.pone.0159722

Liu, L., Cui, H., Fu, R., Zheng, M., Liu, R., Zhao, G., et al. (2017a). The regulation of IMF deposition in pectoralis major of fast-and slow-growing chickens at hatching. *J. Anim. Sci. Biotechnol.* 8 (1), 77. doi: 10.1186/s40104-017-0207-z

Liu, R., Wang, H., Liu, J., Wang, J., Zheng, M., Tan, X., et al. (2017b). Uncovering the embryonic development-related proteome and metabolome signatures in breast muscle and intramuscular fat of fast-and slow-growing chickens. *BMC Genomics* 18 (1), 816. doi: 10.1186/s12864-017-4150-3

Liu, L., Cui, H., Zheng, M., Zhao, G., and Wen, J. (2018). Comparative analysis of differentially expressed genes related to triglyceride metabolism between intramuscular fat and abdominal fat in broilers. *Br. Poult. Sci.* 59 (5), 514–520. doi: 10.1080/00071668.2018.1483573

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2– ΔΔCT method. *Methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Ma, J., Yang, J., Zhou, L., Ren, J., Liu, X., Zhang, H., et al. (2014). A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PloS Genet.* 10 (10), e1004710. doi: 10.1371/journal.pgen.1004710

Ma, Y., Feng, S., Wang, X., Qazi, I. H., Long, K., Luo, Y., et al. (2018). Exploration of exosomal microRNA expression profiles in pigeon 'Milk'during the lactation period. *BMC Genomics* 19 (1), 828. doi: 10.1186/s12864-018-5201-0

Ouyang, H., Wang, Z., Chen, X., Yu, J., Li, Z., and Nie, Q. (2017). Proteomic analysis of chicken skeletal muscle during embryonic development. *Front. Physiol.* 8, 281. doi: 10.3389/fphys.2017.00281

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290. doi: 10.1038/nbt.3122

Pirog, M., Gizak, A., and Rakus, D. (2014). Changes in quaternary structure of muscle fructose-1, 6-bisphosphatase regulate affinity of the enzyme to mitochondria. *Int. J. Biochem. Cell Biol.* 48, 55–59. doi: 10.1016/j.biocel.2013.12.015

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Smith, J. H. (1963). Relation of body size to muscle cell size and number in the chicken. *Poult. Sci.* 42 (2), 283–290. doi: 10.3382/ps.0420283

Sporer, K. R., Tempelman, R. J., Ernst, C. W., Reed, K. M., Velleman, S. G., and Strasburg, G. M. (2011). Transcriptional profiling identifies differentially expressed genes in developing turkey skeletal muscle. *BMC Genomics* 12 (1), 143. doi: 10.1186/1471-2164-12-143

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39 (suppl_2), W316–W322. doi: 10.1093/nar/gkr483

Ylihärsilä, H., Kajantie, E., Osmond, C., Forsen, T., Barker, D., and Eriksson, J. (2007). Birth size, adult body composition and muscle strength in later life. *Int. J. Obesity* 31 (9), 1392. doi: 10.1038/sj.ijo.0803612

Zhan, S., Zhao, W., Song, T., Dong, Y., Guo, J., Cao, J., et al. (2018). Dynamic transcriptomic analysis in hircine longissimus dorsi muscle from fetal to neonatal development stages. *Funct. Integr. Genomics* 18 (1), 43–54. doi: 10.1007/s10142-017-0573-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Lei, Li, Zhou, Gao, Liu, Han and Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 January 2020
doi: 10.3389/fgene.2019.01296

Integrated Hypothalamic Transcriptome Profiling Reveals the Reproductive Roles of mRNAs and miRNAs in Sheep

Zhuangbiao Zhang[1†], Jishun Tang[1,2†], Ran Di[1], Qiuyue Liu[1], Xiangyu Wang[1], Shangquan Gan[3], Xiaosheng Zhang[4], Jinlong Zhang[4], Mingxing Chu[1] and Wenping Hu[1*]*

[1] Key Laboratory of Animal Genetics and Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei, China, [3] State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi, China, [4] Tianjin Institute of Animal Sciences, Tianjin, China

OPEN ACCESS

Edited by:
Robert J. Schaefer, University of Minnesota Twin Cities, United States

Reviewed by:
Ikhide G. Imumorin, Georgia Institute of Technology, United States
Zhibin Ji, Shandong Agricultural University, China

***Correspondence:**
Mingxing Chu
mxchu@263.net
Wenping Hu
huwenping@caas.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *15 May 2019*
Accepted: *25 November 2019*
Published: *15 January 2020*

Citation:
Zhang Z, Tang J, Di R, Liu Q, Wang X, Gan S, Zhang X, Zhang J, Chu M and Hu W (2020) Integrated Hypothalamic Transcriptome Profiling Reveals the Reproductive Roles of mRNAs and miRNAs in Sheep. Front. Genet. 10:1296. doi: 10.3389/fgene.2019.01296

Early studies have provided a wealth of information on the functions of microRNAs (miRNAs). However, less is known regarding their functions in the hypothalamus involved in sheep reproduction. To explore the potential roles of hypothalamic messenger RNAs (mRNAs) and miRNAs in sheep without FecB mutation, in total, 172 and 235 differentially expressed genes (DEGs) and 42 and 79 differentially expressed miRNAs (DE miRNAs) were identified in polytocous sheep in the follicular phase versus monotocous sheep in the follicular phase (PF vs. MF) and polytocous sheep in the luteal phase versus monotocous sheep in the luteal phase (PL vs. ML), respectively, using RNA sequencing. We also identified several key mRNAs (e.g., *POMC*, *GNRH1*, *PRL*, *GH*, *TRH*, and *TTR*) and mRNA–miRNAs pairs (e.g., *TRH* co-regulated by oar-miR-379-5p, oar-miR-30b, oar-miR-152, oar-miR-495-3p, oar-miR-143, oar-miR-106b, oar-miR-218a, oar-miR-148a, and *PRL* regulated by oar-miR-432) through functional enrichment analysis, and the identified mRNAs and miRNAs may function, conceivably, by influencing gonadotropin-releasing hormone (GnRH) activities and nerve cell survival associated with reproductive hormone release *via* direct and indirect ways. This study represents an integral analysis between mRNAs and miRNAs in sheep hypothalamus and provides a valuable resource for elucidating sheep prolificacy.

Keywords: hypothalamus, mRNAs, miRNAs, GnRH, reproduction, sheep

INTRODUCTION

Reproduction, one of the major factors significantly affecting the sheep industry, is a complicated but important physiological process. The success of reproduction is mainly dependent on the release of hormones, including gonadotropin-releasing hormone (GnRH) released from the hypothalamus, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are both secreted from

the pituitary (Cao et al., 2018a). Following the release of hormones, a series of events associated with reproduction, such as ovulation and fertilization, could occur.

It is well known that reproductive traits, such as litter size, are controlled by minor polygene. Researchers have found several major fecundity genes which considerably influence sheep prolificacy, such as bone morphogenetic protein receptor IB (*BMPRIB*), bone morphogenetic protein 15 (*BMP15*) (Chu et al., 2007), and growth differentiation factor 9 (*GDF9*) (Chu et al., 2011). FecB is a mutation in *BMPRIB* occurring in base 746 from A to G. This base change further results in changes in protein function due to a key amino acid transition from glutamine to arginine (Fogarty, 2009). Sheep with one copy of the FecB mutation can experience significant increase in litter size, by 0.67, while this increase is about 1.5 when there are two mutated copies (Liu et al., 2014). Moreover, this mutation was also detected in diverse sheep species, such as Booroola Merino sheep (Mulsant et al., 2001) (Australia), Garole sheep (Polley et al., 2010) (India), Hu sheep (Davis et al., 2006) (China), and Small Tail Han sheep (STH sheep; China) (Davis et al., 2006). STH sheep, an indigenous species in China, has attracted much attention for its excellent traits (Liu et al., 2016; Chao et al., 2017), especially the higher prolificacy (Davis et al., 2006). Furthermore, STH sheep can be divided into three genotypes based on the effects of FecB mutation, better known as FecB BB (with two-copy FecB mutations), FecB B^{+} (with one-copy FecB mutation), and $FecB^{++}$ (with no FecB mutation). Usually, compared to sheep with the other two genotypes, STH sheep with $FecB^{++}$ show a monotocous phenomenon. However, the fact is that there are STH sheep with $FecB^{++}$ and which show a polytocous phenomenon (Davis et al., 2006), and how this mechanism was established remains largely unclear.

With advances in sequencing, the application of RNA sequencing (RNA-seq) in animals, including sheep (Jiang et al., 2014; Zhang et al., 2019a; Zhang et al., 2019b), mice (Beck et al., 2018), and cattle (Correia et al., 2018), enables integral analysis of the expression profiling of mRNA and miRNAs. Therefore, RNA-seq has been widely used to understand some complex traits. Regarding the generation of miRNA, precursor miRNA is transcribed mainly by RNA polymerase II, then processed into mature miRNA (Gebert and Macrae, 2019). Significantly, miRNAs play pivotal roles in life processes, such as muscle growth (Cao et al., 2018c), fleece and hair development (Liu et al., 2018), and neural development (Schratt et al., 2006). Additionally, reproduction is an extremely complex process, and the use of RNA-seq may contribute to enhancing our understanding of sheep fecundity. By comparing the mRNA and miRNA expression patterns in European mouflon and sheep, a research (Yang et al., 2018) found several key mRNAs, such as *INHBA*, *SPP1*, and *ZP2*, and miRNAs, such as miR-374a and miR-9-5p, which may be responsible for the success of female sheep reproduction. Pokharel et al. (2018) detected and characterized some key miRNAs and mRNAs in sheep ovary which may be responsible for sheep prolificacy. Thereby, the identification and functional analysis of mRNAs and miRNAs and characterization of their mutual interaction through sequencing technology may provide new insights into the prolific mechanism in STH sheep with the $FecB^{++}$ genotype, which has so far been difficult to elucidate using standard approaches.

Therefore, in the present study, we applied transcriptomics analysis in PF vs. MF and PL vs. ML to identify DEGs and DE miRNAs and analyze their potential functions, expecting to elucidate the potential prolific mechanism in sheep with the $FecB^{++}$ genotype and act as a reference for other female mammals.

MATERIAL AND METHODS

Preparation of Animals

First, the TaqMan probe (Liu et al., 2017) was applied to genotype the sheep population (n = 890). Then, 12 sheep with no significant differences in sheep age, weight, height, body length, chest circumference, and tube circumference were selected from 142 STH sheep with the $FecB^{++}$ genotype and grouped into the polytocous group (n = 6, litter size ≥2) and monotocous group (n = 6, litter size = 1) according to their litter size records. Additionally, all the sheep were bred under the same conditions, with free access to water and feed, in a sheep farm of the Tianjin Institute of Animal Sciences.

All selected sheep were processed by estrus synchronization with Controlled Internal Drug Releasing Device (CIDR; progesterone 300 mg; Zoetis Australia Pty. Ltd., NSW, Australia) for 12 days. The six sheep, comprising three polytocous sheep and three monotocous sheep, were slaughtered within 45–48 h after CIDR removal (follicular phase), the remaining six sheep were slaughtered on day 9 after CIDR removal (luteal phase). Finally, the selected sheep were divided into four groups, including polytocous sheep in the follicular phase (PF), polytocous sheep in the luteal phase (PL), monotocous sheep in the follicular phase (MF), and monotocous sheep in the luteal phase (ML), on the basis of their littering record and estrous cycle.

Preparation of Tissues, RNA Extraction, and Sequencing

Hypothalamic tissues were collected from 12 killed sheep and immediately stored at −80°C until being used. Then, total RNA was isolated using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) under the manufacturer's instructions, and the quality and integrity of isolated RNA were assessed by an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and electrophoresis. The high-quality RNA of 3 μg of each sample was used to build the mRNA library using a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, USA), which has been described in our previous work (Zhang et al., 2019b). All the sequencing works were conducted in Annoroad Gene Technology Co., Ltd. (Beijing, China).

The fragments with lengths of 18–30 nt, which were obtained from total RNA through the gel separation technique, were used as templates to synthesize the first strand of complementary DNA (cDNA). The second strand of cDNA was also synthesized

in the presence of deoxynucleoside triphosphates (dNTPs), ribonuclease H, and DNA polymerase I. Then the obtained double-stranded cDNA was processed with end-repair, the addition of base A and sequencing adaptors, and uracil-*N*-glycosylase (UNG) enzyme digestion. Finally, polymerase chain reaction was conducted to build the miRNA library.

In addition, a paired-end sequencing approach for mRNAs and miRNAs was conducted using an Illumina HiSeq 2500.

Quality Control, Mapping and Assembly

Raw reads were filtered using in-house software of fqtools_plus-v2.0.0 according to strict criteria, including removing reads with adaptor contaminants, low-quality reads, and reads with N bases accounting for more than 5%. Then, HiSAT2 (Kim et al., 2015) was used to map the cleaned reads to the reference genome (Oarv3.1), and both the sheep reference genome and genome annotation file were downloaded from ENSEMBL (http://www.ensembl.org/index.html). Subsequently, StringTie 1.3.2d (Pertea et al., 2015) was used to assemble transcripts of mRNAs.

Several criteria were also implemented to generate clean miRNA reads, including removing reads without a 3′ adapter, reads without insert fragment, reads with lengths beyond the normal range, raw reads containing too much A/T, and some low-quality reads using in-house scripts. Furthermore, the cleaned data of miRNA were matched against the sheep reference genome (Oarv3.1) by Bowtie v1.1.2 (Langmead et al., 2009).

Differential Expression and Functional Enrichment Analysis of mRNAs

To validate the expression level of mRNAs, the fragments per kilobase per million mapped reads (FPKM) values (Trapnell et al., 2010) were calculated to represent the gene expression level, and DESeq 2-1.4.5 (Wang et al., 2010) was also used to detect the DEGs between two comparisons based on FPKM values. Additionally, a gene with fold change >1.5 and $p < 0.05$ was considered as a DEG in PF vs. MF and PL vs. ML. In addition, we also performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. We first downloaded the Uniprot database, where each sequence contains the GO annotation and KEGG annotation species (sheep) of the sequence as well as gene and protein names. All genes of sheep to be analyzed were compared with the Uniprot database by blast (NCBI-blast 2.2.28) to find the best alignment result for each sequence, and corresponding to GO and KEGG annotation results. Then, we also downloaded the corresponding relationship between the entry name and number provided on the websites of GO and KEGG, as well as the classification hierarchy file, and summarize the GO and KEGG classification of the genes we obtained. Lastly, a particular GO term or KEGG pathway with a hypergeometric p value < 0.05 was thought to indicate significant enrichment.

Differential Expression Analysis and Prediction of Target Genes of miRNAs

The miRDeep v2.0.0.8 (Friedländer et al., 2012) was applied to identify the known and novel miRNAs by mapping clean reads and hairpins to mature miRNAs recorded in the miRbase database (Griffiths-Jones, 2006). In addition, transcripts per million (TPM) were calculated to represent miRNA expression levels on the basis of the reads number. DESeq2-1.4.5 (Wang et al., 2010) was also applied to identify DE miRNAs in PF vs. MF and PL vs. ML, and the threshold of fold change >1.5, $p < 0.05$ was considered to indicate differential expression. Furthermore, miRanda v3.3a (Enright et al., 2004) was used to predict the target genes of miRNAs.

Integral miRNA–mRNA Networks Analysis

To precisely identify key DE miRNAs and DEGs associated with reproduction, a network containing DE miRNAs and DE mRNAs, on the basis of miRNA functions (Gebert and Macrae, 2019), was built using Cytoscape_v3.5.0 (Shannon et al., 2003), and only mRNAs exhibiting negative relationship with miRNAs were included in miRNA–mRNA interaction networks.

Data Validation

In order to validate the accuracy of sequencing data, four DEGs, including *CRH*, *FOXG1*, *TTR*, and *POMC*, and four DE miRNAs, including oar-miR-433-3p, oar-miR-495-3p, oar-miRNA-16b, and oar-miR-143, were selected for data validation. First, the primers of DEGs and DE miRNAs were synthesized by Beijing Tianyi Huiyuan Biotechnology Co., Ltd. (Beijing, China) (**Supplementary Table 1**) for subsequent reverse transcription, which was performed using PrimeScript™ RT reagent kit (TaKaRa) for mRNAs and miRcute Plus miRNAs First-Strand cDNA Kit (TIANGEN, Beijing, China) for miRNA. Furthermore, quantitative PCR (qPCR) was conducted with the SYBR Green qPCR Mix (TaKaRa, Dalian, China) for mRNAs and miRcute Plus miRNA qPCR Kit (TIANGEN, Beijing, China) for miRNAs using a RocheLight Cycler®480 II system (Roche Applied Science, Mannheim, Germany). In addition, β-actin (for mRNA) and U6 small nuclear RNA (snRNA; for miRNA) were utilized as reference gene/miRNA to calculate the relative expression level with the method of $2^{-\Delta\Delta ct}$ (Livak and Schmittgen, 2001). The qPCR for mRNAs was conducted in the following procedure: initial denaturation at 95°C for 5 minutes, followed by 40 cycles of denaturation at 95°C for 5 s, then annealing at 60°C for 30 s. While the qPCR for miRNA was conducted in the following procedure: initial denaturation at 95°C for 15 minutes, followed by 40 cycles of denaturation at 94°C for 20 s, then annealing at 60°C for 34 s. All the qPCR results were presented as the mean ± SD.

RESULTS

mRNA and miRNA Profiling

To fully characterize the globally hypothalamic mRNA and miRNA expression differences between sheep with the same genotype but different litter sizes, RNA-seq was used to detect their expression profile in the hypothalamus. In total, RNA-seq for mRNA generated approximately 1,519 million raw reads and 1,460 million clean reads (**Supplementary Table 2**) after

data filtering. Overall, 21,221 mRNAs were identified (**Supplementary Table 3**) after mapping to sheep genome, and our results also suggested that many mRNAs were located in the intergenic region (nearly 45%), followed by the intron (about 35%) and exon (more than 20%) regions (**Figure 1A** and **Supplementary Table 4**).

Regarding the expression level of mRNAs, our results showed that the FPKM of those genes obtained from RNA-seq at <50 constituted nearly 90%, and the high-expression genes, i.e., those with FPKM >500, constituted about 0.5% (**Supplementary Table 3**), which suggested that the data obtained from the hypothalamus *via* RNA-seq were relatively reasonable. Furthermore, the chromosome distribution of mRNAs indicated that chromosome 3 contains 9.79% of the genes identified from the hypothalamus, followed by chromosome 1 (9.55%) and chromosome 2 (7.22%) (**Figure 1B** and **Supplementary Table 5**). Additionally, the number of DEGs identified from PF vs. MF (**Figure 2A** and **Supplementary Table 6**) and PL vs. ML (**Figure 2B** and **Supplementary Table 6**) were 172 and 235, respectively. Among these DEGs, 79 and 90 were upregulated, while 93 and 145 were downregulated in PF vs. MF and PL vs. ML, respectively. In addition, the expression density of DEGs displayed obviously different expression patterns between PF and MF, and between PL and ML (**Figures 2C**, **D**).

Regarding miRNAs, RNA-seq generated approximately 315 million raw reads and 267 million clean reads (**Supplementary Table 7**) with lengths ranging from 18 to 30 nt (**Figure 3A**) after removing low-quality reads. Overall, 623 miRNAs were detected (**Supplementary Table 8**). In addition, the chromosome distribution of identified miRNAs was also determined. As **Figure 3B** shows, the chromosome distribution of miRNAs from 1 to *X* varies (**Supplementary Table 9**), and most of the identified miRNAs were located at chromosome 3 (nearly 40%),

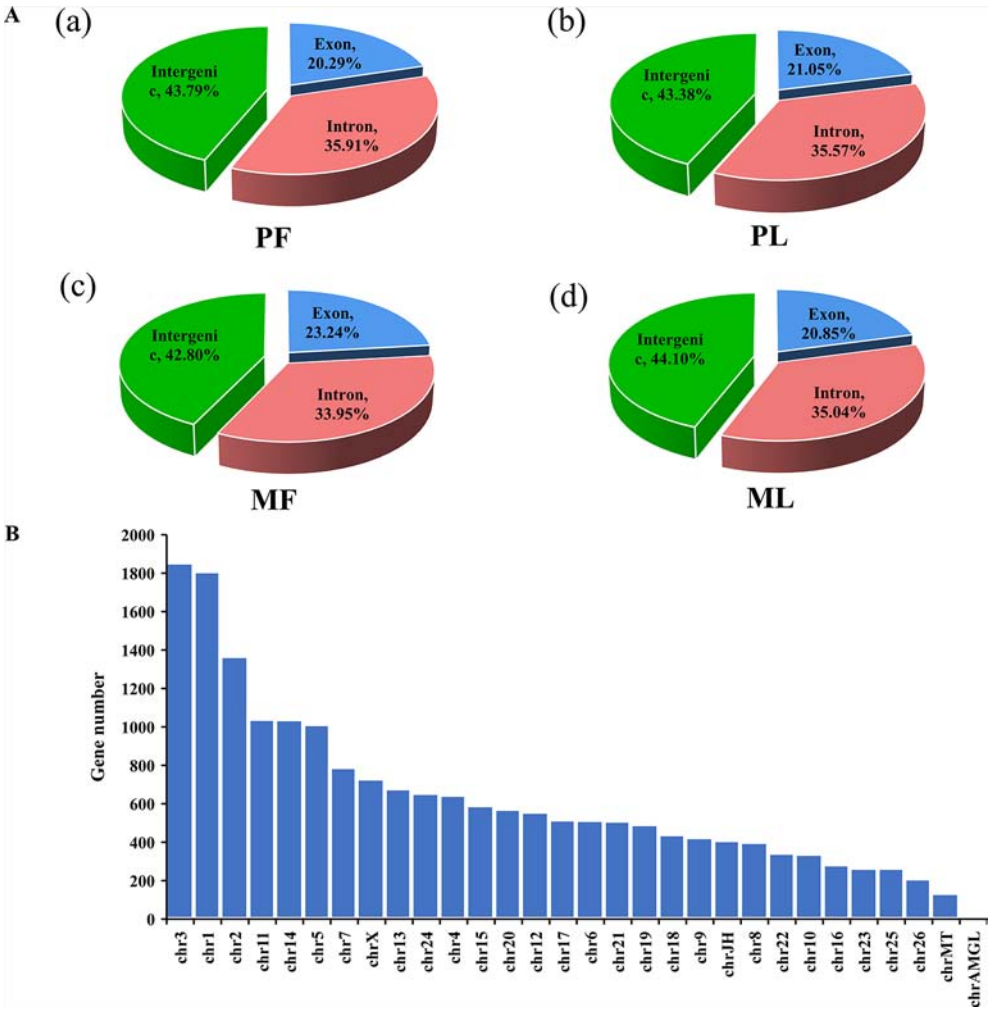


FIGURE 1 | Mapping region and chromosome distribution of identified mRNAs. **(A)** Mapping region of identified genes at the reference genome in polytocous sheep in the follicular phase (PF) (**a**), polytocous sheep in the luteal phase (PL) (**b**), monotocous sheep in the follicular phase (MF) (**c**), and monotocous sheep in the luteal phase (ML) (**d**). **(B)** Chromosome distribution of identified genes from the hypothalami in the PF, MF, PL, and ML.

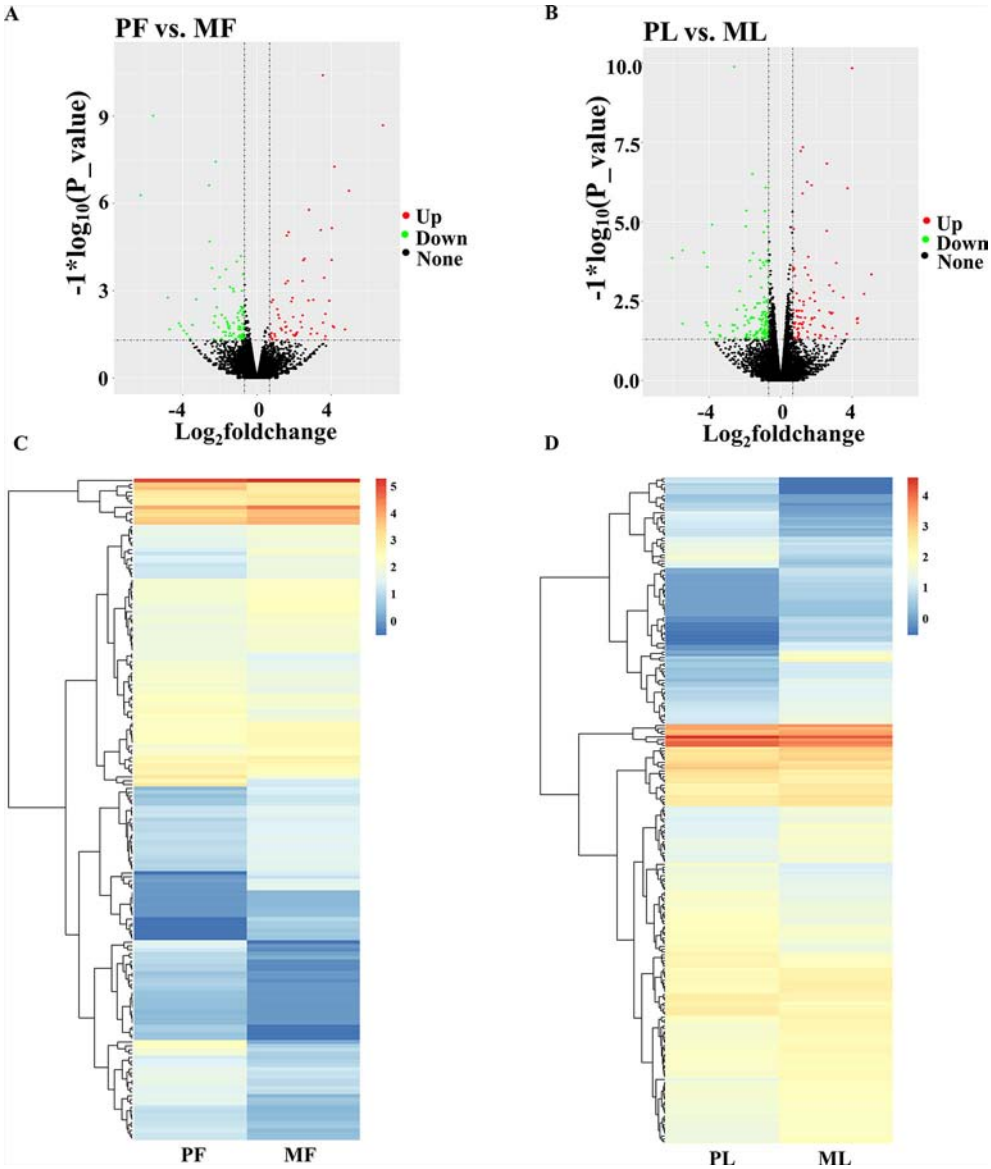


FIGURE 2 | Differentially expressed genes (DEGs) analysis. **(A)** Volcano plot of identified genes in PF vs. MF, where *red* and *green* represent up- or downregulation, respectively, same below. **(B)** Volcano plot of identified genes in PL vs. ML. **(C)** Heat maps showing the expression intensity of 794 DEGs in the follicular phase, including PF and MF. **(D)** Heat maps showing the expression intensity of 1,044 DEGs in the luteal phase, including PL and ML.

followed by chromosome 9 (nearly 15%) and chromosome 18 (nearly 9%). Interestingly, chromosome 3 also contains the most mRNAs (**Figure 3B**). Also, a diversity of non-coding RNAs (ncRNAs), including transfer RNAs (tRNAs), snRNAs, miRNAs, etc., were also identified (**Figure 3C** and **Supplementary Table 10**), and the known miRNAs account only for a small part of all the identified ncRNAs. In addition, the target genes of miRNAs in PF vs. MF and PL vs. ML were predicted to be 1,611 and 2,120, respectively (**Supplementary Table 11**).

Additionally, the DE miRNAs identified from PF vs. MF and PL vs. ML were 42 and 79, respectively. Of these DE miRNAs, 20 and 23 were upregulated, while 22 and 56 were downregulated, respectively (**Figure 4A** and **Supplementary Table 12**). In addition, the expression density of DEGs displayed obviously different expression patterns between PF and MF, and between PL and ML (**Figures 4B, C**).

GO and KEGG Enrichment Analysis of DEGs

To better understand the potential functions of the DEGs, GO term and KEGG pathway analyses were performed. In GO analysis, the most enriched term in PF vs. MF was the MHC

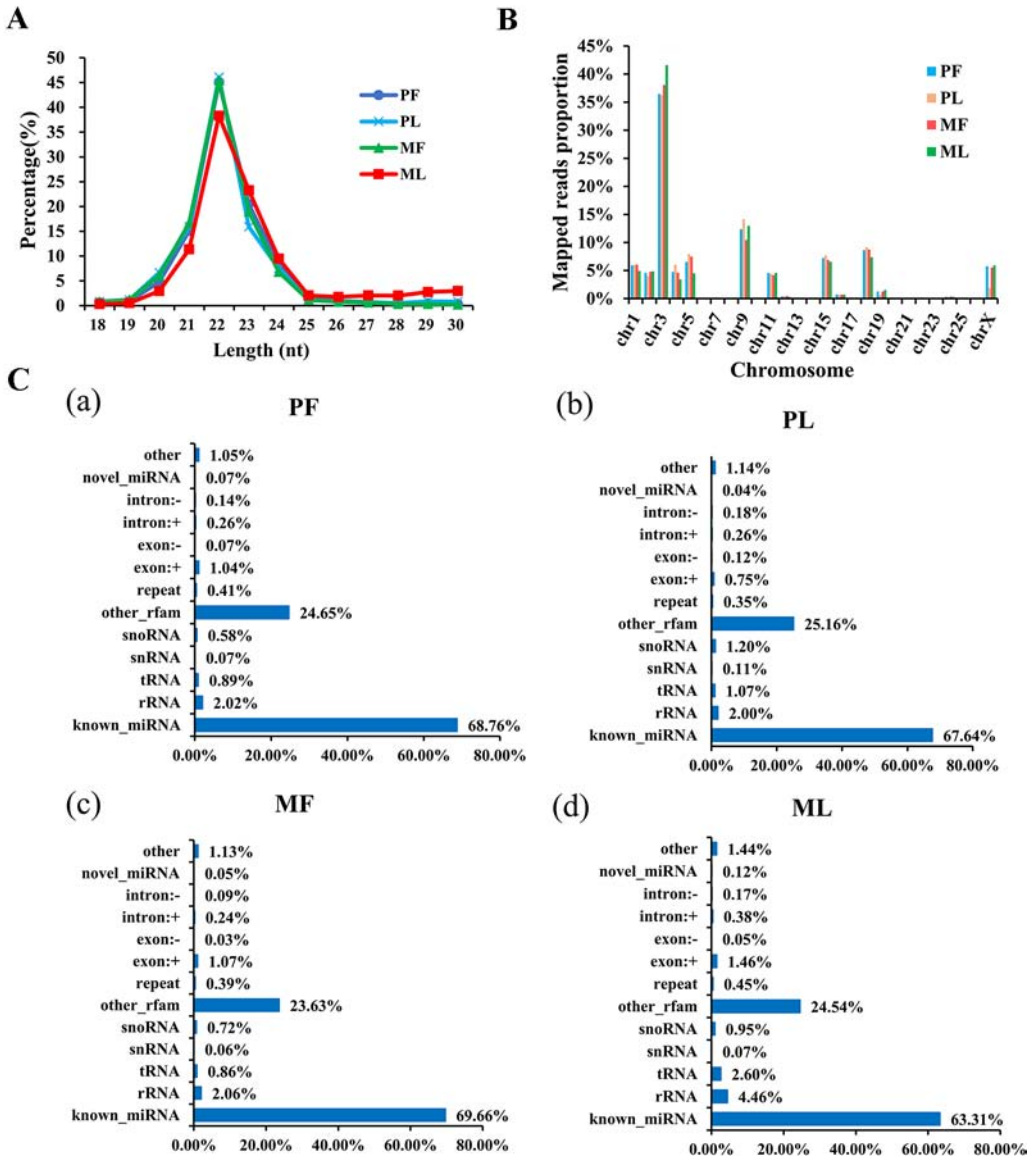


FIGURE 3 | Characterization of microRNA (miRNA) profiling and the percentage of detected miRNAs from ncRNAs. **(A)** Length distribution of clean reads from identified miRNA fragments. **(B)** The chromosome distribution of identified miRNAs from hypothalami. **(C)** Categories of identified non-coding RNAs (ncRNAs) *via* sequencing in PF (**a**), PL (**b**), MF (**c**), and ML (**d**).

protein complex (GO:0042611). Other GO terms related to the MHC protein were also enriched, such as MHC class II protein complex binding (GO:0023026) and MHC protein complex binding (GO:0023023), indicating the crucial role of the MHC protein in the hypothalamic functions (**Figure 5A** and **Supplementary Table 13**). Regarding PL vs. ML, the top 2 enriched terms were the immune system process (GO:0002376) and immune response (GO:0006955). In addition, some GO terms associated with chemokine receptors, including CXCR3 chemokine receptor binding(GO:0048248) and chemokine receptor binding (GO:0042379), were also highly enriched, suggesting the important roles of the immune system and chemokine receptors in the hypothalamus at the luteal phase (**Figure 5A** and **Supplementary Table 13**).

KEGG analysis in PF vs. MF (**Figure 5B** and **Supplementary Table 14**) showed that the most enriched pathway was type I diabetes mellitus (map04940). In addition, other metabolic pathways, such as alpha-linolenic acid metabolism (map00592) and arachidonic acid metabolism(map00590), were also enriched. Regarding PL vs. ML, the top enriched pathways were cytokine–cytokine receptor interaction (map04060). A pathway named the Jak-STAT signaling pathway (map04630), which has been found to participate in the reproductive process (Ko et al., 2018), was also enriched.

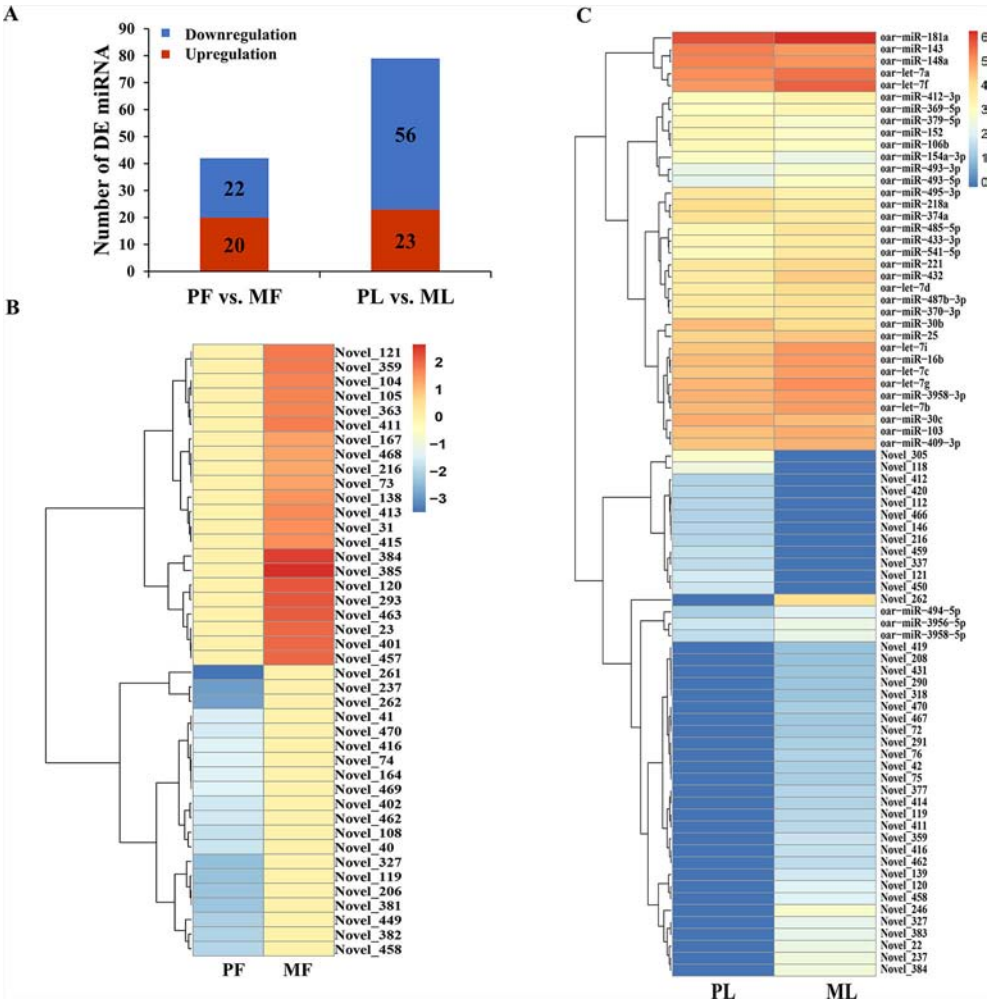


FIGURE 4 | Differentially expressed (DE) microRNA (miRNA) analysis. **(A)** DE miRNAs in PF vs. MF and PL vs. ML. Heat maps showing the expression intensity of 42 and 79 DE miRNAs in the follicular phase including PF and MF **(B)** and the luteal phase including PL and ML **(C)**, the names of miRNAs were also labeled.

Analysis of Integrated miRNA–mRNA Co-Expression Network

To fully understand the potential reproductive roles of miRNAs, we built interactome networks using DE miRNAs and their targets (DEGs). In total, 42 DE miRNAs (novel miRNAs) in PF vs. MF were predicted to target 1,611 genes (**Supplementary Table 15**). The number of overlapped genes, which means the target genes were also DEGs, was 8 (**Figure 6A** and **Supplementary Table 16**). An mRNA–miRNA co-expression network was then constructed, where 5 DEGs were targeted by 3 novel miRNAs (**Figure 6B**). Regarding PL vs. ML, 38 known and 41 novel DE miRNAs were predicted to target 1,747 and 1,659 genes (**Supplementary Table 15**), and the numbers of overlapped genes were 179 and 9, respectively (**Figures 6C, D** and **Supplementary Table 16**). The main upregulated miRNA–mRNA co-expression network suggested that 55 DEGs were targeted by 11 DE miRNAs containing the top 10 upregulated known miRNAs and one novel miRNA (**Figure 6E**). The main downregulated miRNA–mRNA co-expression network suggested that 33 DEGs were targeted by 11 DE miRNAs containing the top 10 downregulated known miRNAs and one novel miRNA (**Figure 6F**).

Data Validation

>In order to assess the accuracy of sequencing, qPCR was applied to verify the RNA-seq data. The results indicated that both mRNAs and miRNAs in sheep hypothalamus displayed expression patterns similar to the sequencing results (**Figure 7**), demonstrating the reliability of the data generated from RNA-seq.

DISCUSSION

In this study, we initially identified 172 and 235 DEGs, and 42 and 79 DE miRNAs in two comparisons (PF vs. MF and PL vs.

FIGURE 5 | Functional enrichment analysis of DEGs. **(A)** Top enriched GO terms at the biological process, molecular function, and cellular component level in PF vs. MF and PL vs. ML, in addition, the gray represents no enrichment, same below. **(B)** Top enriched KEGG pathways in PF vs. MF and PL vs. ML.

ML) through RNA-seq. Of these DE miRNAs, miRNA family members including the let-7 and oar-miRNA-200 family exhibited differential expression levels. Furthermore, one study detecting 48 DE miRNAs from sheep ovary, including the let-7 and oar-miRNA-200 family members, suggested that those identified miRNAs were differentially expressed in seasonal and non-seasonal sheep breeds (Zhai et al., 2018). Therefore, some miRNAs, such as let-7 and oar-miRNA-200 family members may not be only species-specific but also phase- or fecundity-specific in sheep. In addition, some miRNAs, including miRNA-138 and miRNA-212, were detected in rat hypothalamus (Amar et al., 2012), which differed significantly from miRNAs identified in sheep hypothalamus (both miRNA-138 and miRNA-212 in our results failed to be detected). Besides,

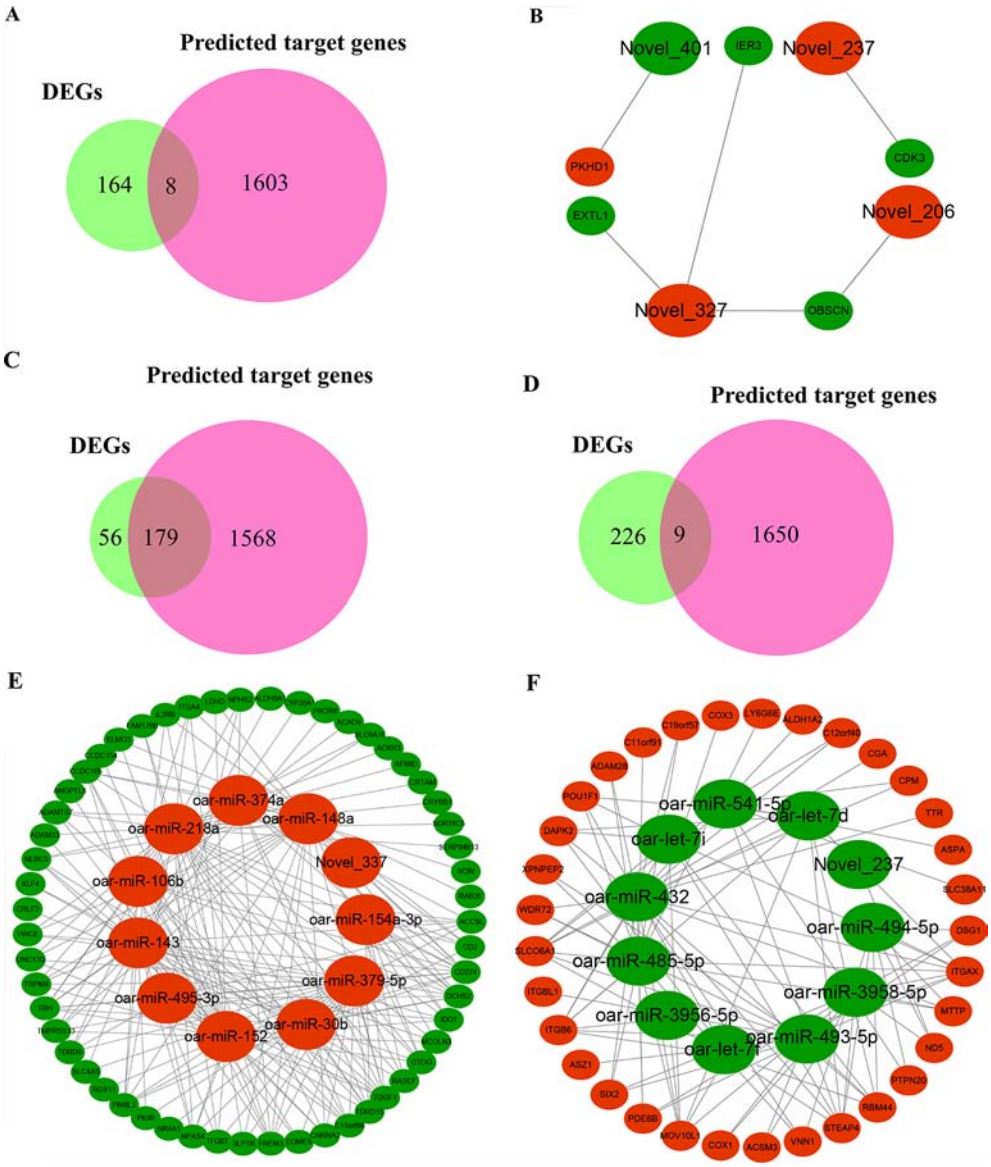


FIGURE 6 | Overview of mRNA–miRNA networks. **(A)** Overlapped genes in PF vs. MF between DEGs and miRNA-targeted genes. **(B)** Hypothalamus network in PF vs. MF of four miRNAs containing three upregulated miRNAs and one downregulated miRNA and five target genes, where *large* and *small ellipses* represent miRNAs and DEGs, respectively, in addition, *red* and *green* represent up- or downregulation, respectively; same below. **(C)** Overlapped genes in PL vs. ML between DEGs and predicted target genes negatively modulated by known miRNAs. **(D)** Overlapped genes in PL vs. ML between DEGs and novel miRNA-targeted genes. **(E)** Main hypothalamic upregulated network in PL vs. ML containing the top 10 upregulated miRNAs and one novel miRNA and 55 target genes. **(F)** Main hypothalamic downregulated network in PL vs. ML containing the top 10 downregulated miRNAs and one novel miRNA and 33 target genes.

several miRNAs, such as miRNA-200 family members, were conserved in the hypothalamus of mice (Choi et al., 2008; Crépin et al., 2014), rat (Sangiaoalvarellos et al., 2014), and zebrafish (Garaffo et al., 2015), as well as sheep (our results). In summary, we confirmed that several miRNAs are conserved in many animals, but there were also miRNAs that showed a species-specific distribution in the hypothalamus, which means those differences may be responsible for the differences between sheep and rats, and even other non-mammals.

Functional Analysis of DEGs in PF vs. MF

In the functional enrichment analysis of DEGs in PF vs. MF, several key genes, including prolactin (*PRL*), proopiomelanocortin (*POMC*), and gonadotropin releasing hormone 1 (*GNRH1*), were found to participate in the reproductive process. Some researchers have proven that PRL and E2 could respond rapidly to stimulation in the arcuate nucleus (ARC) of rat hypothalamic slices (Nishihara and Kimura, 1989). Araujo-Lopes et al. (2014) revealed that PRL could regulate the activities of GnRH through modulating kisspeptin

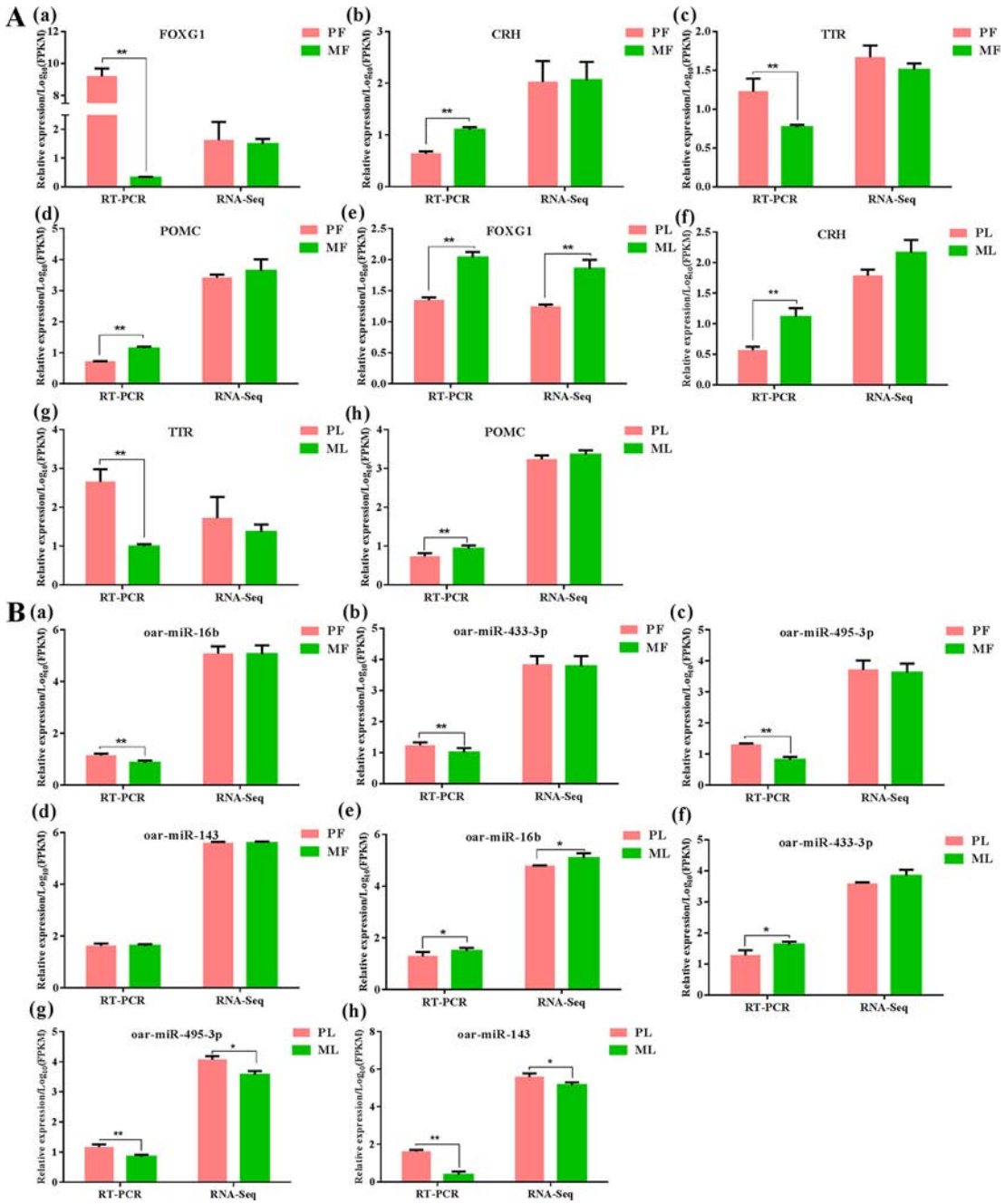


FIGURE 7 | Data validation of mRNAs **(A)** and miRNAs **(B)** by qPCR in PF, PL, MF and ML, meanwhile ** represents $p < 0.01$, while * represents $p < 0.05$. *FOXG1*: Forkhead box L1, *CRH*: corticotropin-releasing hormone, *TTR*: transthyretin, *POMC* proopiomelanocortin.

neurons in the ARC of female rats and inhibit LH secretion, causing a series of alterations in the estrous cycle. Our results indicated that the expression of *PRL* in PF was more than three times that of PRL in MF. Therefore, coupled with the inhibitory role of PRL on LH, we speculate that *PRL* may affect LH or FSH activities by influencing the pulsatile GnRH wave in the hypothalamus.

POMC neurons, as a key upstream factor affecting hypothalamic hormone release, were found to be sensitive to metabolic hormones such as leptin (Wilson and Enriori, 2015) and enhance kisspeptin neuron activities in rodents, resulting in increased GnRH secretion (Muroi and Ishii, 2016). Leptin can act in the hypothalamus directly, eliciting the release of GnRH (Guzmán et al., 2019), and promoting the expression of *POMC* (Perello et al., 2007). Although the stimulatory effects of POMC on kisspeptin have been known for a long time, how this signaling is established remains poorly understood (Saedi et al., 2018). Significantly, our results indicated

that the expression of *POMC* in PF was relatively lower than in MF, while *GNRH1*, which has been reported to play a key role in determining sheep litter size (An et al., 2013), displayed a reverse expression pattern between PF and MF. Therefore, we hypothesized that a negative regulatory relationship between POMC and GNRH1 may exist in sheep hypothalamus.

Functional Analysis of DEGs in PL vs. ML

In functional enrichment analysis of DEGs in PL vs. ML, some pathways including the Jak-STAT signaling pathway (*PRL*, *GH*, *CRLF2*, ENSOARG00000007618, ENSOARG00000016231, and *IL2RB*) were highly enriched. The current study argued that the Jak-STAT signaling pathway in mice was involved in GnRH activities (Ko et al., 2018). *PRL*, as mentioned above, plays an important role in GnRH activities (Araujo-Lopes et al., 2014). The expression of *PRL* was detected not only in the follicular phase but also in the luteal phase, and interestingly, there was a reverse expression pattern of *PRL* between PF vs. MF and PL vs. ML, suggesting its crucial roles in reproduction. The effects of leptin on GnRH release have been revealed (Guzmán et al., 2019), and the infusion of leptin into the arcuate nucleus in rats could cause *PRL* release (Watanobe, 2010), which suggested that *PRL* can be a downstream factor activated by leptin to function in GnRH activities. In addition, the overexpression of growth hormone (GH) could disrupt the state of reproduction, mainly through mediating leptin activities (Chen et al., 2018). Additionally, estrogen could play an inhibitory role on GH *in vivo* (Leung et al., 2003). Collectively, considering the effects of *PRL* and GH on leptin, we speculated that GH, leptin, and *PRL* may coordinate to inhibit GnRH release.

The Regulatory Network of miRNA–mRNA After Transcription in PF vs. MF

To better understand the functions of miRNAs, a negative interactome containing 5 mRNAs and 4 miRNAs in PF vs. MF was built. Cyclin-dependent kinase 3 (*CDK3*), targeted by Novel_237, was reported that the downregulation of activities of CDK3-related kinase could promote cell apoptosis in the rat (Braun et al., 1998). Immediate early response 3 (*IER3*), targeted by Novel_327, was also involved in enhancing (Zhou et al., 2017) or mediating (Jin et al., 2015) cell apoptosis. Polycystic kidney and hepatic disease gene 1 (*PKHD1*), targeted by Novel_401, has been discovered to induce cell apoptosis, after being downregulated through the PI3K and NF-κB pathways (Sun et al., 2011). Furthermore, our sequencing data indicated that *CDK3* and *IER3* were downregulated while *PKHD1* was upregulated in PF vs. MF. All in all, we hypothesized that more nerve cell apoptosis occurred in MF than PF, which may further influence hormone activities associated with reproduction and may lead to the final observed litter size differences.

The Regulatory Network of miRNA–mRNA After Transcription in PL vs. ML

The regulatory network of miRNA–mRNA after transcription in PL vs. ML was divided into two main negative networks: the main upregulated and the main downregulated network. In the main upregulated network, thyrotropin-releasing hormone (*TRH*), co-regulated by oar-miR-379-5p, oar-miR-30b, oar-miR-152, oar-miR-495-3p, oar-miR-143, oar-miR-106b, oar-miR-218a, and oar-miR-148a, has been reported to function in GnRH release (see below). Triclosan in mice was found to reduce the production of TRH and thyroid-stimulating hormone (TSH), and this decreased effect could further cause hyperprolactinemia. Hyperprolactinemia was suggested to cause a suppressive effect on kisspeptin expression, resulting in deficits in reproductive and endocrine function (Cao et al., 2018b). In addition, TRH can not only stimulate *PRL* release but also inhibit LH release, and this inhibitory effects may occur through prohibiting the release of GnRH (Araujo-Lopes et al., 2014). Collectively, TRH in the hypothalamus may be responsible, at least in part, for the suppression of GnRH activities.

In the main downregulated network of miRNAs, transthyretin (*TTR*) was reversely regulated by oar-miR-432. The expression level of *TTR*'s in rats could be enhanced by progesterone *via* progesterone receptors both *in vitro* and *in vivo* (Quintela et al., 2011), and a similar upregulated effect of TTR caused by progesterone in mouse uterus was also observed (Diao et al., 2010). Furthermore, TTR could drive the nuclear translocation of insulin-like growth factor 1 receptor (IGF-1R) (Vieira et al., 2015), which could lead to functional changes in insulin-like growth factor 1 (IGF1). Interestingly, the stimulatory effect of IGF1 on GnRH release has been discovered (Hiney et al., 2009). Therefore, we speculated that the negative feedback effects of progesterone on GnRH release may be mediated by TTR, which reduces the binding probability between IGF1 and its receptor, further resulting in a suppression of GnRH activities.

All results indicated that several key DEGs and DE miRNAs in the hypothalamus directly or indirectly participate in hormone activities associated with reproduction, and further studies involving gene/miRNA knockout or overexpression could help us to understand their real functions in female reproductive traits.

CONCLUSION

As far as we know, this study provides the first integral mRNA–miRNA interactome in sheep without FecB mutation from the perspective of the hypothalamus. We identified several DEGs (e.g., *POMC*, *GNRH1*, *PRL*, *TRH*, and *TTR*) and mRNA–miRNA pairs (e.g., *TRH* coagulated by oar-miR-379-5p, oar-miR-30b, oar-miR-152, oar-miR-495-3p, oar-miR-143, oar-miR-106b, oar-miR-218a and oar-miR-148a and *PRL* regulated by oar-miR-432) from the RNA-seq data obtained from sheep hypothalamus, which may function through influencing the activities of GnRH. Our results provide novel insights into the prolificacy mechanism of sheep, which may facilitate the discovery of novel major genes and a deeper understanding of female sheep reproduction.

DATA AVAILABILITY STATEMENT

All the data obtained from RNA-seq has been deposited in the Sequence Read Archive database under the bioproject numbers PRJNA529384 and PRJNA532808.

ETHICS STATEMENT

The animal study was reviewed and approved by the Science Research Department (in charge of animal welfare issues) of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (IAS-CAAS) (Beijing, China). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

WH and MC designed the research. ZZ wrote the paper. JT, ZZ, WH, SG, XZ, and JZ collected the data. ZZ performed the study. ZZ and JT analyzed data. MC and WH revised the final manuscript. All authors reviewed the manuscript and approved the final version.

FUNDING

This work was supported by the National Natural Science Foundation of China (31501941, 31772580 and 31472078), the Genetically Modified Organisms Breeding Major Program of China (2016ZX08009-003-006 and 2016ZX08010-005-003), the Earmarked Fund for China Agriculture Research System (CARS-38), the Central Public-Interest Scientific Institution Basal Research Fund (2018-YWF-YB-1, Y2017JC24, 2017ywf-zd-13), the Agricultural Science and Technology Innovation Program of China (ASTIP-IAS13, CAAS-XTCX2016010-01-03, CAAS-XTCX2016010-03-03, CAAS-XTCX2016011-02-02), the China Agricultural Scientific Research Outstanding Talents and Their Innovative Teams Program, the China High-level Talents Special Support Plan Scientific and Technological Innovation Leading Talents Program (W02020274), and the Tianjin Agricultural Science and Technology Achievements Transformation and Popularization Program (201704020), Joint Funds of the National Natural Science Foundation of China and the Government of Xinjiang Uygur Autonomous Region of China (U1130302). The APC was funded by the National Natural Science Foundation of China (31772580).

ACKNOWLEDGMENTS

We thank Annoroad Gene Technology (Beijing) Co., Ltd. for RNA-sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01296/full#supplementary-material

REFERENCES

Amar, L., Benoit, C., Beaumont, G., Vacher, C. M., Crepin, D., and Taouis, M. (2012). MicroRNA expression profiling of hypothalamic arcuate and paraventricular nuclei from single rats using Illumina sequencing technology. *J. Neurosci. Methods* 209, 134–143. doi: 10.1016/j.jneumeth.2012.05.033

An, X. P., Hou, J. X., Zhao, H. B., Li, G., Bai, L., Peng, J. Y., et al. (2013). Polymorphism identification in goat *GNRH1* and *GDF9* genes and their association analysis with litter size. *Anim. Genet.* 44, 234–238. doi: 10.1111/j.1365-2052.2012.02394.x

Araujo-Lopes, R., Crampton, J. R., Aquino, N. S., Miranda, R. M., Kokay, I. C., Reis, A. M., et al. (2014). Prolactin regulates kisspeptin neurons in the arcuate nucleus to suppress LH secretion in female rats. *Endocrinology* 155, 1010–1020. doi: 10.1210/en.2013-1889

Beck, A. S., Wood, T. G., Widen, S. G., Thompson, J. K., and Barrett, A. D. T. (2018). Analysis by deep sequencing of discontinued neurotropic yellow fever vaccine strains. *Sci. Rep.* 8, 13408. doi: 10.1038/s41598-018-31085-2

Braun, K., Hölzl, G., Pusch, O., and Hengstschläger, M. (1998). Deregulated expression of CDK2- or CDK3-associated kinase activities enhances c-Myc-induced apoptosis. *DNA Cell Biol.* 17, 789–798. doi: 10.1089/dna.1998.17.789

Cao, C., Ding, Y., Kong, X., Feng, G., Xiang, W., Chen, L., et al. (2018a). Reproductive role of miRNA in the hypothalamic-pituitary axis. *Mol. Cell. Neurosci.* 88, 130–137. doi: 10.1016/j.mcn.2018.01.008

Cao, X. Y., Hua, X., Xiong, J. W., Zhu, W. T., Zhang, J., and Chen, L. (2018b). Impact of triclosan on female reproduction through reducing thyroid hormones to suppress hypothalamic kisspeptin neurons in mice. *Front. Mol. Neurosci.* 11, 6. doi: 10.3389/fnmol.2018.00006

Cao, Y., You, S., Yao, Y., Liu, Z. J., Hazi, W., Li, C. Y., et al. (2018c). Expression profiles of circular RNAs in sheep skeletal muscle. *Asian-australas. J. Anim. Sci.* 31, 1550–1557. doi: 10.5713/ajas.170563

Chao, T., Wang, G., Ji, Z., Liu, Z., Hou, L., Wang, J. M., et al. (2017). Transcriptome analysis of three sheep intestinal regions reveals key pathways and hub regulatory genes of large intestinal lipid metabolism. *Sci. Rep.* 7, 5345. doi: 10.1038/s41598-017-05551-2

Chen, J., Cao, M., Zhang, A., Shi, M., Tao, B., Li, Y., et al. (2018). Growth hormone overexpression disrupts reproductive status through actions on leptin. *Front. Endocrinol.* 9, 131. doi: 10.3389/fendo.2018.00131

Choi, P. S., Zakhary, L., Choi, W. Y., Caron, S., Alvarez-Saavedra, E., Miska, E. A., et al. (2008). Members of the miRNA-200 family regulate olfactory neurogenesis. *Neuron* 57, 41–55. doi: 10.1016/j.neuron.2007.11.018

Chu, M. X., Liu, Z. H., Jiao, C. L., He, Y. Q., Fang, L., Ye, S. C., et al. (2007). Mutations in *BMPR-IB* and *BMP-15* genes are associated with litter size in Small Tailed Han sheep (*Ovis aries*). *J. Anim. Sci.* 85, 598–603. doi: 10.2527/jas.2006-324

Chu, M. X., Yang, J., Feng, T., Cao, G. L., Fang, L., Di, R., et al. (2011). *GDF9* as a candidate gene for prolificacy of Small Tail Han sheep. *Mol. Biol. Rep.* 38, 5199–5204. doi: 10.1007/s11033-010-0670-5

Correia, C. N., McLoughlin, K. E., Nalpas, N. C., Magee, D. A., Browne, J. A., Rue-Albrecht, K., et al. (2018). RNA Sequencing (RNA-Seq) reveals extremely low levels of reticulocyte-derived globin gene transcripts in peripheral blood from horses (*Equus caballus*)) and cattle (*Bos taurus*). *Front. Genet.* 9, 278. doi: 10.3389/fgene.2018.00278

Crépin, D., Benomar, Y., Riffault, L., Amine, H., Gertler, A., and Taouis, M. (2014). The over-expression of miR-200a in the hypothalamus of ob/ob mice is linked to leptin and insulin signaling impairment. *Mol. Cell. Neurosci.* 384, 1–11. doi: 10.1016/j.mce.2013.12.016

Davis, G. H., Balakrishnan, L., Ross, I. K., Wilson, T., Galloway, S. M., Lumsden, B. M., et al. (2006). Investigation of the Booroola (*FecB*) and Inverdale ($FecX^{(I)}$) mutations in 21 prolific breeds and strains of sheep sampled in 13 countries. *Anim. Reprod. Sci.* 92, 87–96. doi: 10.1016/j.anireprosci.2005.06.001

Diao, H., Xiao, S., Cui, J., Chun, J., Xu, Y., and Ye, X. (2010). Progesterone receptor-mediated up-regulation of transthyretin in preimplantation mouse uterus. *Fertil. Steril.* 93, 2750–2753. doi: 10.1016/j.fertnstert.2010.01.009

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2004). MicroRNA targets in Drosophila. *Genome Biol.* 5, R1. doi: 10.1186/gb-2003-5-1-r1

Forgaty, N. M. (2009). A review of the effects of the Booroola gene (FecB) on sheep production. *Small Ruminant Res.* 85, 75–84. doi: 10.1016/j.smallrumres.2009.08.003

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Garaffo, G., Conte, D., Provero, P., Tomaiuolo, D., Luo, Z., Pinciroli, P., et al. (2015). The Dlx5 and FOXG1 transcription factors, linked *via* miRNA-9 and -200, are required for the development of the olfactory and GnRH system. *Mol. Cell. Neurosci.* 68, 103–119. doi: 10.1016/j.mcn.2015.04.007

Gebert, L. F. R., and Macrae, I. J. (2019). Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20, 21–37. doi: 10.1038/s41580-018-0045-7

Griffiths-Jones, S. (2006). MiRBase : the MicroRNA Sequence Database. *Methods Mol. Biol.* 342, 129–138. doi: 10.1385/1-59745-123-1:129

Guzmán, A., Hernández-Coronado, C. G., Rosales-Torres, A. M., and Hernández-Medrano, J. H. (2019). Leptin regulates neuropeptides associated with food intake and GnRH secretion. *Ann. Endocrinol.* 80, 38–46. doi: 10.1016/j.ando.2018.07.012

Hiney, J. K., Srivastava, V. K., Pine, M. D., and Les Dees, W. (2009). Insulin-like growth factor-I activates KISS-1 gene expression in the brain of the prepubertal female rat. *Endocrinology* 150, 376–384. doi: 10.1210/en.2008-0954

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173. doi: 10.1126/science.1252806

Jin, H., Suh, D. S., Kim, T. H., Yeom, J. H., Lee, K., and Bae, J. (2015). IER3 is a crucial mediator of TAp73β-induced apoptosis in cervical cancer and confers etoposide sensitivity. *Sci. Rep.* 5, 8367. doi: 10.1038/srep08367

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth3317

Ko, E. K., Chorich, L. P., Sullivan, M. E., Cameron, R. S., and Layman, L. C. (2018). JAK/STAT signaling pathway gene expression is reduced following Nelf knockdown in GnRH neurons. *Mol. Cell. Endocrinol.* 470, 151–159. doi: 10.1016/j.mce.2017.10.009

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25

Leung, K. C., Doyle, N., Ballesteros, M., Sjogren, K., Watts, C. K., Low, T. H., et al. (2003). Estrogen inhibits GH signaling by suppressing GH-induced JAK2 phosphorylation, an effect mediated by SOCS-2. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1016–1021. doi: 10.1073/pnas.0337600100

Liu, Q., Pan, Z., Wang, X., Hu, W., Di, R., Yao, Y., et al. (2014). Progress on major genes for high fecundity in ewes. *Front. Agric. Sci. Eng* 1, 282–290. doi: 10.15302/J-FASE-2014042

Liu, Z. Z., Ji, Z. B., Wang, G. Z., Chao, T. L., Hou, L., and Wang., J. M. (2016). Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics* 17, 863. doi: 10.1186/s12864-016-3212-2

Liu, Q., Hu, W., He, X., Pan, Z., Guo, X., Feng, T., et al. (2017). Establishment of high-throughput molecular detection methods for ovine high fecundity major gene *FecB* and their application. *Acta Veterinaria Et Zootechnica Sin.* 48, 39–51. doi: 10.11843/j.issn.0366-6694.2017.01.005

Liu, Y., Zhang, J., Xu, Q., Kang, X., Wang, K., Wu, K., et al. (2018). Integrated miRNA-mRNA analysis reveals regulatory pathways underlying the curly fleece trait in Chinese tan sheep. *BMC Genomics* 19, 360. doi: 10.1186/s12864-018-4736-4

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta ct}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., et al. (2001). Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5104–5109. doi: 10.1073/pnas.091577598

Muroi, Y., and Ishii, T. (2016). A novel neuropeptide Y neuronal pathway linking energy state and reproductive behavior. *Neuropeptides* 59, 1–8. doi: 10.1016/j.npep.2016.09.002

Nishihara, M., and Kimura, F. (1989). Postsynaptic effects of prolactin and estrogen on arcuate neurons in rat hypothalamic slices. *Neuroendocrinology* 49, 215–218. doi: 10.1159/000125117

Perello, M., Stuart, R. C., and Nillni, E. A. (2007). Differential effects of fasting and leptin on proopiomelanocortin peptides in the arcuate nucleus and in the nucleus of the solitary tract. *Am. J. Physiol. Endocrinol. Metab.* 292, 300–313. doi: 10.1152/ajpendo.004662006

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt3122

Pokharel, K., Peippo, J., Honkatukia, M., Seppälä, A., Rautiainen, J., Ghanem, N., et al. (2018). Integrated ovarian mRNA and miRNA transcriptome profiling characterizes the genetic basis of prolificacy traits in sheep (*Ovis aries*). *BMC Genomics* 19, 104. doi: 10.1186/s12864-017-4400-4

Polley, S., De, S., Brahma, B., Mukherjee, A., Vinesh, P. V., Batabyal, S., et al. (2010). Polymorphism of *BMPR1B*, *BMP15* and *GDF9* fecundity genes in prolific Garole sheep. *Trop. Anim. Health Prod.* 42, 985–993. doi: 10.1007/s11250-009-9518-1

Quintela, T., Gonçalves, I., Martinho, A., Alves, C. H., Saraiva, M. J., Rocha, P., et al. (2011). Progesterone enhances transthyretin expression in the rat choroid plexus *in vitro* and *in vivo via* progesterone receptor. *J. Mol. Neurosci.* 44, 152–158. doi: 10.1007/s12031-010-9398-x

Saedi, S., Khoradmehr, A., Mohammad Reza., J. S., and Tamadon, A. (2018). The role of neuropeptides and neurotransmitters on kisspeptin/kiss1r-signaling in female reproduction. *J. Chem. Neuroanat.* 92, 71–82. doi: 10.1016/j.jchemneu.2018.07.001

Sangiaoalvarellos, S., Penabello, L., Manfredilozano, M., Tenasempere, M., and Cordido, F. (2014). Perturbation of hypothalamic microRNA expression patterns in male rats after metabolic distress: impact of obesity and conditions of negative energy balance. *Endocrinology* 155, 1838–1850. doi: 10.1210/en.2013-1770

Schratt, G. M., Tuebing, F., Nigh, E. A., Kane, C. G., Sabatini, M. E., Kiebler, M., et al. (2006). A brain-specific microRNA regulates dendritic spine development. *Nature* 439, 283–289. doi: 10.1038/nature04367

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., and Ramage, D. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sun, L., Wang, S., Hu, C., and Zhang, X. (2011). Down-regulation of *PKHD1* induces cell apoptosis through PI3K and NF-κB pathways. *Exp. Cell. Res.* 317, 932–940. doi: 10.1016/j.yexcr.2011.01.025

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt1621

Vieira., M., Gomes, J. R., and Saraiva, M. J. (2015). Transthyretin induces insulin-like growth factor i nuclear translocation regulating its levels in the hippocampus. *Mol. Neurobiol.* 51, 1468–1479. doi: 10.1007/s12035-014-8824-4

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Watanobe, H. (2010). Leptin directly acts within the hypothalamus to stimulate gonadotropin-releasing hormone secretion *in vivo* in rats. *J. Physiol.* 545, 255–268. doi: 10.1113/jphysiol.2002.023895

Wilson, J. L., and Enriori, P. J. (2015). A talk between fat tissue, gut, pancreas and brain to control body weight. *Mol. Cell. Endocrinol.* 418, 108–119. doi: 10.1016/j.mce.2015.08.022

Yang., J., Li, X., Cao, Y. H., Pokharel, K., Hu, X. J., Chen, Z. H., et al. (2018). Comparative mRNA and miRNA expression in European mouflon (*Ovis musimon*) and sheep (*Ovis aries*) provides novel insights into the genetic mechanisms for female reproductive success. *Heredity* 122, 172–186. doi: 10.1038/s41437-018-0090-1

Zhai, M., Xie, Y., Liang, H., Lei, X., and Zhao, Z. (2018). Comparative profiling of differentially expressed microRNAs in estrous ovaries of Kazakh sheep in different seasons. *Gene* 664, 181–191. doi: 10.1016/j.gene.2018.04.025

Zhang, Z. B., Tang, J. S., Di, R., Liu, Q. Y., Wang, X. Y., Gan, S. Q., et al. (2019a). Comparative transcriptomics reveal key sheep (*Ovis aries*) hypothalamus lncRNAs that affect reproduction. *Animals* 9, 15. doi: 10.3390/ani9040152

Zhang, Z. B., Tang, J. S., He, X. Y., Zhu, M. X., Gan, S. Q., Guo, X. F., et al. (2019b). Comparative transcriptomics identify key hypothalamic circular RNAs that participate in Sheep (*Ovis aries*) reproduction. *Animals* 9, 557. doi: 10.3390/ani9080557

Zhou, Q., Hahn, J. K., Neupane, B., Aidery, P., Labeit, S., Gawaz, M., et al. (2017). Dysregulated IER3 expression is associated with enhanced apoptosis in titin-based dilated cardiomyopathy. *Int. J. Mol. Sci.* 18, 723. doi: 10.3390/ijms1804723

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhang, Tang, Di, Liu, Wang, Gan, Zhang, Zhang, Chu and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 January 2020
doi: 10.3389/fgene.2019.01339

Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs

Brittney N. Keel, Warren M. Snelling, Amanda K. Lindholm-Perry, William T. Oliver, Larry A. Kuehn and Gary A. Rohrer*

USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE, United States

OPEN ACCESS

Edited by:
David E. MacHugh, University College Dublin, Ireland

Reviewed by:
Fabyano Fonseca Silva, Universidade Federal de Viçosa, Brazil
Eduardo Casas, National Animal Disease Center (USDA ARS), United States

***Correspondence:**
Brittney N. Keel
brittney.keel@ars.usda.gov

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *24 July 2019*
Accepted: *09 December 2019*
Published: *21 January 2020*

Citation:
Keel BN, Snelling WM, Lindholm-Perry AK, Oliver WT, Kuehn LA and Rohrer GA (2020) Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs. Front. Genet. 10:1339. doi: 10.3389/fgene.2019.01339

The "large *p* small *n*" problem has posed a significant challenge in the analysis and interpretation of genome-wide association studies (GWAS). The use of prior information to rank genomic regions and perform SNP selection could increase the power of GWAS. In this study, we propose the use of gene expression data from RNA-Seq of multiple tissues as prior information to assign weights to SNP, select SNP based on a weight threshold, and utilize weighted hypothesis testing to conduct a GWAS. RNA-Seq libraries from hypothalamus, duodenum, ileum, and jejunum tissue of 30 pigs with divergent feed efficiency phenotypes were sequenced, and a three-way gene x individual x tissue clustering analysis was performed, using constrained tensor decomposition, to obtain a total of 10 gene expression modules. Loading values from each gene module were used to assign weights to 49,691 commercial SNP markers, and SNP were selected using a weight threshold, resulting in 10 SNP sets ranging in size from 101 to 955 markers. Weighted GWAS for feed intake in 4,200 pigs was performed separately for each of the 10 SNP sets. A total of 36 unique significant SNP associations were identified across the ten gene modules (SNP sets). For comparison, a standard unweighted GWAS using all 49,691 SNP was performed, and only 2 SNP were significant. None of the SNP from the unweighted analysis resided in known QTL related to swine feed efficiency (feed intake, average daily gain, and feed conversion ratio) compared to 29 (80.6%) in the weighted analyses, with 9 SNP residing in feed intake QTL. These results suggest that the heritability of feed intake is driven by many SNP that individually do not attain genome-wide significance in GWAS. Hence, the proposed procedure for prioritizing SNP based on gene expression data across multiple tissues provides a promising approach for improving the power of GWAS.

Keywords: constrained tensor decomposition, gene expression, clustering, feed efficiency, swine, GWAS, weighted SNP

INTRODUCTION

The "large *p* small *n*" problem has posed a significant challenge in the analysis and interpretation of genome-wide association studies (GWAS; Diao and Vidyashankar, 2013). The problem refers to the scenario in statistical inference where the dimension of independent variables, *p*, is larger than the sample size, *n*. Typically in GWAS, the number of observations, *n*, is in the hundreds or thousands and the number of markers, *p*, is in the hundreds of thousands. Statistical procedures such as shrinkage estimation and variable selection are often employed to ensure the existence solutions in *large-p-small-n* regressions in GWAS (Fernando et al., 2017).

The most commonly used approach to GWAS is single-SNP analysis, where linear or logistic regression is performed separately for each SNP followed by multiple-testing correction. This standard single-step adjustment disregards prior knowledge of potentially noteworthy regions, and, as a result, tests of significance for SNP in such regions may be overly down-weighted due to the other relatively inconsequential SNP. Hence, using prior information to rank genomic regions and perform SNP selection could increase the power of GWAS.

Recent advances in statistical methodology have made it possible to incorporate prior information through weighted hypothesis testing (Genovese et al., 2006). Roeder et al. (2006) introduced a method which uses linkage analysis information to up- or down-weight SNP according to their prior likelihood of association with a trait of interest, and the resulting weighted P-values are used in the false discovery rate (FDR) procedure. A similar approach using expression quantitative trait loci (eQTL) information to weight SNP was proposed by Li et al. (2013).

Transcriptome sequencing (RNA-Seq) is a widely used technology for genome-wide transcript quantification, used to analyze gene expression patterns, and provide insight into the mechanisms underlying complex traits in livestock species. Genome-wide gene expression data from thousands of studies have been accumulating and made available through public repositories such as the Gene Expression Omnibus (GEO; Edgar et al., 2002). Recently, GWAS results have been interpreted by interrogating significant SNP for associations with gene expression data in livestock (Ballester et al., 2017; Fang et al., 2017; Kommadath et al., 2017; Cai et al., 2018; Deng et al., 2019). These studies have integrated GWAS and gene expression data post-GWAS. In this study, we propose the use of gene expression data from RNA-Seq of multiple tissues (hypothalamus, duodenum, ileum, and jejunum) as prior information to assign weights to SNP, select SNP based on a weight threshold, and utilize weighted hypothesis testing to conduct a GWAS for swine feed efficiency.

MATERIAL AND METHODS

The U.S. Meat Animal Research Center (USMARC) Animal Care and Use committee reviewed and approved the use of animals in this study.

Population

Feed intake and body weight gain were measured on cohorts of growing pigs reared at USMARC. All pigs were sired by either Landrace or Yorkshire boars sourced from 5 different genetic suppliers and produced out of Landrace-Yorkshire cross sows. Two different genetic suppliers are represented in each group of pigs. Pigs entered the barn at approximately 95 days of age at the beginning of the feeding trial and had *ad libitum* access to a standard corn/soybean meal-based diet that met or exceeded NRC requirements (NRC, 2012). Pigs in each cohort (196 per cohort) were assigned to one of 14 same-sex pens (14 pigs per pen) containing a single Feed Intake Recording System (FIRE) feeder (Osborne Industries, Inc., Osborne KS). After a 1-week adjustment period, daily feed intakes for each pig were recorded *via* the FIRE feeders and pigs were weighed at the beginning (d0) and end (d 42) of the feeding trial. Twenty-two cohorts of pigs had individual feeding events recorded.

Different numbers of animals from the population were used in different stages of the study. Feed intake data was collected on a total of 4,200 animals across the 22 cohorts. Four of these 22 cohorts (n = 784 animals) were used to select 30 animals with extreme feed efficiency phenotypes for RNA-Seq. Lastly, GWAS was performed using data from the 2,813 animals that were both genotyped and phenotyped. Detailed descriptions of each stage of the study are provided in subsequent sections.

Sampling for RNA-Seq

Feed efficiency phenotypes were determined for each pig in four cohorts (n = 784 animals) by dividing average daily body weight gain (ADG) by average daily feed intake (ADFI) to determine the gain to feed ratio (Gain : Feed). From each cohort of pigs, a selection criterion was applied to select animals for further study that included ADG within ± 0.30 SD of the mean and the greatest and least ADFI (n = 7 or 8 per cohort). The descriptive statistics are presented in **Table 1**.

Tissue Collection, RNA Isolation, and Sequencing

Tissue collection and RNA extraction were performed using the same procedures in each contemporary group. Sample collection time frame was consistent across cohorts. Pigs identified as high and low feed efficiency were euthanized with barbiturates in accordance with the American Veterinary Medical Association guidelines (AVMA, 2013). The head was removed, and the hypothalamus was collected and stored at -80°C as previously described (Thorson et al., 2017). One 3-cm segment of mid-jejunum and one 3-cm segment of mid-ileum were collected from pigs as previously described (Oliver et al., 2002). In addition, a 3-cm segment of duodenum was collected approximately 5-cm caudal of the cranial duodenal flexure.

Total RNA was isolated from the tissue samples using the RNeasy Mini Plus kit and QiaShredder columns (Qiagen, Valenci, CA, USA). Briefly, 800 ul of RLT buffer with β-mercaptoethanol were added to 50–100 mg of tissue samples and homogenized for 40 sec using an Omni Prep 6-station homogenizer (Omni International, Kennesaw, GA, USA). The

TABLE 1 | Descriptive data of efficient and inefficient pigs (n = 30).[1,2]

	ADFI[3], kg/d	Initial Weight, kg	Ending Weight, kg	ADG[4], kg/d	Gain : Feed
Efficient (Low Intake)	2.08 ± 0.11	47.7 ± 2.8	88.8 ± 2.8	0.991 ± 0.04	0.458 ± 0.025
Inefficient (High Intake)	2.80 ± 0.11	51.1 ± 3.0	93.6 ± 3.2	1.025 ± 0.05	0.367 ± 0.16

[1]*Animals selected included those with ADG within ± 0.30 SD of the mean and the greatest (inefficient) and least (efficient) ADFI.*
[2]*Data means ± SEM.*
[3]*Average daily feed intake.*
[4]*Average daily gain.*

homogenate was centrifuged in a QiaShredder column on full speed for 3 min. Genomic DNA was removed from the total RNA with the Qiagen RNeasy Plus mini-kit, according to the manufacturer's protocol, and the total RNA was eluted in 50 ul of RNase free water. Total RNA was quantified with a NanoDrop One spectrophotometer (Thermo Scientific, Wilmington, DE). The average 260/280 ratio was 2.05, with a range of 1.94–2.09. An Agilent Bioanalyzer RNA 6000 nano kit (Santa Clara, CA, USA) was used to determine the RNA integrity number (RIN). Only samples with a RIN of 8.0 and higher were used for the RNA sequencing. The average RIN was 9.1, with a range of 8.1–9.9.

Samples were prepared for RNA sequencing with the Illumina TruSeq Stranded mRNA High Throughput Sample kit and protocol (Illumina Inc., San Diego, CA, USA). The libraries were quantified with qRT-PCR using the NEBNext Library Quant Kit (New England Biolabs, Inc., Beverly, MA, USA) on a CFX384 thermal cycler (Bio-Rad, Hercules, CA, USA), and the quality of the library was determined with an Agilent Bioanalyzer DNA 1000 kit (Santa Clara, CA, USA). The libraries were diluted with Tris-HCL 10 mM, pH 8.5 with 0.1% Tween 20 to 4nM samples (Teknova, Hollister, CA. USA). All libraries were paired-end sequenced with 150 cycle high output sequencing kits for the Illumina NextSeq instrument. Bases of the paired-end reads for all sequenced libraries were identified with the Illumina BaseCaller, and FASTQ files were produced for downstream analysis of the sequence data. Sequence data is available for download from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) BioProjects PRJNA528599 (hypothalamus), PRJNA528884 (duodenum), PRJNA529214 (ileum), and PRJNA529662 (jejunum).

Sequence Data Processing

Read alignment of the RNA-Seq reads was carried out as follows. First, quality of the raw paired-end sequence reads in individual FASTQ files was assessed using FastQC (Version 0.11.5; www.bioinformatics.babraham.ac.uk/projects/fastqc), and reads were trimmed to remove adapter sequences and low-quality bases using the Trimmomatic software (Version 0.35; Bolger et al., 2014). The remaining reads were mapped to the Sscrofa 11.1 genome assembly using Hisat2 (Version 2.1.1; Kim et al., 2015) with its default parameters. The StringTie software (Pertea et al., 2015) was then used to calculate raw read counts for each of the 29,651 annotated genes in the NCBI Sscrofa 11.1 reference annotation (Release 106).

Filtering of lowly expressed genes and normalization of read counts was performed using a protocol that considers the multi-tissue structure of the data. First, raw read counts were normalized using the median of ratios normalization scheme from the DESeq2 software package (Love et al., 2014), where read counts are divided by sample-specific size factors determined by median ratio of gene counts relative to the geometric mean per gene. A normalized gene expression matrix was constructed for each tissue, and the arithmetic mean of expression values across samples within each tissue was computed. Genes with mean normalized expression < 100 in all 6 tissues were removed from further analysis.

Three-Way Clustering *Via* Constrained Tensor Decomposition to Detect Gene Expression Modules

Three-way clustering of multi-tissue, multi-individual gene expression data was performed using an adaption of the method described by Wang et al. (2017). Gene expression measurements for the four tissues were organized into a 3-way array, or order-3 tensor, with gene, individual, and tissue modes. That is, the input to the algorithm was an order-3 tensor given by, $\Omega = [\![\omega ijk]\!] \in \mathbb{R}^{n_G \times n_I \times n_T}$, where ω_{ijk}, denotes the normalized gene expression value for gene i in individual j in tissue k, n_G the number of genes, n_I the number of individuals, and n_T the number of tissues. The tensor Ω was then decomposed into a sum R of rank-1 components,

$$\Omega = \sum_{r=1}^{R} \lambda_r \boldsymbol{G}_r \otimes \boldsymbol{I}_r \otimes \boldsymbol{T}_r + \varepsilon, \quad (1)$$

where $\lambda_1 \geq \lambda_1 \geq \ldots \geq \lambda_R \geq 0$ are singular values in decreasing order, and $\boldsymbol{G}_r$, $\boldsymbol{I}_r$, and $\boldsymbol{T}_r$ are norm-1 singular vectors that indicate the relative contribution of each gene, individual, and tissue to the r-th component, respectively, and $\varepsilon = [E_{ik}]$ is a noise tensor with each entry E_{ik} i.i.d. $N(0,\sigma^2)$.

Complete details of the algorithm used for tensor decomposition can be found in Wang et al. (2017). Briefly, the successive rank-1 approximation to Ω is determined by iteratively solving the following minimization problem:

$$\underset{\lambda_r,\ \boldsymbol{G}_r,\boldsymbol{I}_r,\boldsymbol{T}_r}{\text{minimize}} \| \Omega - \lambda_r \boldsymbol{G}_r \otimes \boldsymbol{I}_r \otimes \boldsymbol{T}_r \|_F, \quad (2)$$

$$\text{subject to } \boldsymbol{G}_r \|_2 = \| \boldsymbol{I}_r \|_2 = \| \boldsymbol{T}_r \|_2 = 1,$$

where $\|\cdot\|_F$ is defined entry-wise as $\Omega_F = \sqrt{\sum_{i=1}^{n_G}\sum_{i=1}^{n_I}\sum_{i=1}^{n_T}\omega_{ijk}^2}$. At each iteration, we imposed one of two conditions, either $\boldsymbol{T}_r \geq 0$ or $\boldsymbol{T}_r \leq 0$, by thresholding values in to 0. The appropriate sign of was selected to maximize $\boldsymbol{T}_r$. This constraint on eases the interpretation of the interaction at the tissue level. Non-zero tissue loading values indicate that the module is "active" in the

tissue. Without constraining values in to a single sign, it is possible (in fact likely) that contains two expression modules, one for the tissues with positive loading values and one for the tissues with negative loading values. Consequently, gene and individual loadings become less informative since they cannot be explicitly assigned to either the positive or negative loading module. Note that the constraint on used in this work is slightly different that of Wang et al. (2017), where they imposed strict non-negativity on $\boldsymbol{T}_r$.

Genes with large values in $\boldsymbol{G}_r$ exhibit strong relationships with individuals and tissues in the r-th component, while these relationships are stronger in the individuals with larger $\boldsymbol{I}_r$-values and tissues with larger (in absolute value) $\boldsymbol{T}_r$-values. The loading vectors $\boldsymbol{G}_r$, $\boldsymbol{I}_r$, and $\boldsymbol{T}_r$ will be referred to as eigen-genes, eigen-individuals, and eigen-tissues, respectively, throughout the remainder of the manuscript.

Gene Ontology Enrichment Analysis

Enrichment analysis of gene ontology GO terms was performed using the PANTHER classification system (Version 14.1; Mi et al., 2016). PANTHER's implementation of the binomial test of overrepresentation with the default Ensembl *Sus scrofa* GO annotation as background was utilized. Data from PANTHER was considered statistically significant at FDR-corrected $P \leq 0.05$.

Characterization of Gene Expression Modules

GO enrichment analysis was performed on the top genes within each expression module, where the top genes in module r were defined as genes having a loading value in $\boldsymbol{G}_r$ greater than a specified cutoff value, c, which controls the significance level. A permutation-based approach was used to determine c with an arbitrarily selected significance level of $\alpha = 0.005$. One hundred null tensors were generated by randomly and independently permuting gene expression values for every individual-tissue pair. That is,

$$\boldsymbol{\Omega}^{\text{null}}(\mathcal{G}, \text{ individual } j, \text{ tissue } k) \stackrel{\text{def}}{=} \boldsymbol{\Omega}(\mathcal{PG}, \text{ individual } j, \text{ tissue } k),$$

where $\mathcal{G}$ denotes the original set of gene expression values and $\mathcal{PG}$ denotes the permutated gene expression values. Each null tensor was decomposed, and their eigen-genes were used to represent the null distribution of gene expression values within each module. The cutoff value for module r, c_r, was the 99.5-percentile of the empirical distribution of $\boldsymbol{G}_r^{\text{null}}$.

Proportion of Variance in Individual Loadings

Sources of variation in individual loadings were analyzed by fitting the following linear model:

$$I_j = \beta_1 + \beta_2 \text{ADFI}_j + \beta_3 \text{CG}_j + \beta_4 \text{Gender}_j + \varepsilon_j, \quad (3)$$

where ADFI denotes average daily feed intake, CG denotes contemporary group, $\boldsymbol{I} = (I_1, \ldots, I_{n_I})^T$, and $\varepsilon_j \sim N(0, \sigma^2)$ for all $j = 1, 2, \ldots, n_I$. After the model was fit, the proportion of variance explained by each covariate (ADFI, contemporary group, and gender) was calculated using ANOVA.

Tensor Projection for Identifying ADFI-Associated Genes

Using the notation from above, let $\boldsymbol{\Omega} \in \mathbb{R}^{n_G \times n_I \times n_T}$ denote the expression tensor and $\{\boldsymbol{T_r} \in \mathbb{R}^{n_T}\}$ be the set of eigen-tissues from the tensor decomposition. Let $\boldsymbol{\Omega}(\cdot, \cdot, \boldsymbol{T_r})$ be the tensor projection of $\boldsymbol{\Omega}$ through the eigen-tissue $\boldsymbol{T_r} = (T_{r,1}, \ldots, T_{r,n_T})^T$, i.e.,

$$\boldsymbol{\Omega}(\cdot, \cdot, \boldsymbol{T_r}) = \sum_{k=1}^{n_T} T_{r,k} \, \boldsymbol{\Omega}(\cdot, \cdot, k) \, . \quad (4)$$

Then, the following linear model was used for each gene tested,

$$\begin{aligned} &\boldsymbol{\Omega}(\text{test gene}, \ \cdot, \boldsymbol{T_r}) \\ &\quad = \beta_1 \boldsymbol{1} + \beta_2 \mathbf{ADFI} + \beta_3 \mathbf{CG} + \beta_4 \mathbf{Gender} + \boldsymbol{\varepsilon}, \end{aligned} \quad (5)$$

where $\varepsilon_j \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$. The ADFI-effect was assessed by testing $\mathcal{H}_0 : \beta_2 = 0$ against $\mathcal{H}_\alpha : \beta_2 \neq 0$.

Phenotypic Data Collection for Genetic Association Analysis

Twenty-two cohorts of 196 pigs had individual feeding events recorded in a building fitted with Osbourne FIRE Feeders. The animals and facilities were previously described in Section 2.1. Records were removed for animals with incomplete data due to one of the following reasons: animal removed from the study due to health, failure of the electronic ID eartag, or failure of the FIRE Feeder for a majority of the test. As a result, 4,200 animals remained in the study. Aberrant feeding events were removed if they did not conform to a logical length of meal time (1 sec < meal time < 3,600 sec), amount of feed consumed (20 g < feed consumed < 3 kg), and consumption rate (rate < 2 kg/min). Once aberrant feeding events were removed, feeding parameters were computed for each pen and day of test to determine if a feeder was not operating properly. Statistics used to remove a pen x day included number of aberrant feeding events recorded, amount of feed distributed, and total number of events for each day. After all suspicious records were removed, the amount of feed consumed by each pig for each day of test was calculated, resulting in a total of 164,660 records of the 184,800 possible daily intake records.

Data were analyzed with WOMBAT (Version 17-07-2017; Meyer, 2007) fitting a random regression mixed model. Fixed effects fitted were gender (barrow or gilt) and a combined group-pen effect. Day on test was fit as the independent variable using a cubic Legendre polynomial, and animal was fitted as a random effect. A cubic Legendre polynomial was selected as it dramatically improved the log likelihood of the model over a quadratic Legendre regression and only marginal improvements were seen when evaluating higher order polynomials. Random regression coefficients were projected to individual daily intake for each of the days on test, to fill the missing intake records and

adjust for fixed effects. Daily projections were summed to obtain adjusted test intake for each individual.

Genotypic Data Collection for Genetic Association Analysis

Tail samples were collected on all pigs and stored at −20°C. Genomic DNA was extracted using the WIZARD genomic DNA purification kit according to the manufacturer's protocol (Promega Corp., Madison, WI, USA). Genotyping was conducted using three platforms: the NeoGen Porcine GGPHD chip (GeneSeek, Lincoln, USA), Illumina Porcine SNP60 v2 chip (Illumina, Inc., San Diego, USA), and NeoGen GGP-Porcine chip (GeneSeek, Lincoln, USA).

Genetic Association Analysis

Ancestors of the pigs having intake records were identified from USMARC pedigree records to create a 7,009 animal pedigree. Phenotyped pigs and their ancestors genotyped with a SNP assay, Illumina Porcine SNP60 v1 or v2 (Illumina, Inc., San Diego, USA), Illlumina Porcine SNP50 (Illumina, Inc., San Diego, USA), NeoGen GGP-Porcine chip (GeneSeek, Lincoln, USA), and NeoGen Porcine GGPHD chip (GeneSeek, Lincoln, USA) were identified. The SNP were ordered according to the Sscrofa11.1 genome assembly and available pedigree was used to impute genotypes to 49,695 SNP from at least one assay for the 4,632 genotyped animals (2,813 phenotyped, 1,819 ancestors) using findhap (VanRaden et al., 2013).

Following VanRaden (2008), weighted genomic relationship matrices (G), were constructed as

$$G = \frac{M^{\star\prime} M^{\star}}{2\Sigma_{i=1}^{m} p_i(1-p_i)}, \tag{6}$$

where m is the number of SNP, p_i the frequency of the B allele for the i^{th} SNP, and M* a centered genotype matrix (M) weighted by a diagonal matrix of weighting factors (D)

$$M^{*} = MD\,. \tag{7}$$

Genomic relationship matrices were constructed for equally weighted SNP (D = m x m identity matrix) as well as for gene-centric weightings. Weights for SNP within gene boundaries were calculated as $-log_{10}(P)$, where P denotes the P-value obtained from testing the ADFI-effect in Equation (5) in the gene module of interest. If a SNP did not reside in a gene, it was assigned a weight of zero.

For a given weight threshold, t, three G for each of the 10 sets of gene weightings were evaluated: (1) a weighted analysis with all SNP where all SNP had non-zero weightings (min = 0.00001), (2) an unweighted analysis using only SNP with weight > t, and (3) a weighted analysis using only SNP with weight > t. Arbitrary thresholds of $t = 2$ and $t = 5$ were evaluated.

The average information restricted maximum likelihood (AIREML) algorithm implemented in WOMBAT was used to estimate heritability (h^2) of test intake attributable to pedigree relationships and each weighted genomic relationship matrix. Phenotypic variance should remain constant; all estimates of phenotypic variance from these data using different unweighted G were similar. Weighted G resulted in additive variance estimates much greater than phenotypic variance from unweighted G, and residual variances were similar to estimates using unweighted G. Assuming the residual variance estimate is appropriate for variation not explained by weighted G and phenotypic variance equal to that estimated with unweighted G, the amount of variation explained by weighted G should be the difference between phenotypic variance from unweighted G and residual variance from weighted G, and corrected heritability that difference divided by phenotypic variance. That is,

$$h_w^2 = \frac{Var(P_u) - Var(E_w)}{Var(P_u)}$$

where P_u denotes the phenotypic variance from unweighted G, and E_w denotes the residual variance from weighted G.

After convergence, effects of individual SNP were estimated for each genomic relationship matrix. Following Wang et al. (2012),

$$\hat{a} = M^{\star\prime}\left[M^{\star}M^{\star\prime}\right]^{-1}\hat{u_g}, \tag{8}$$

where $\hat{a}$ is a vector of SNP effect estimates and $\hat{u_g}$ the vector of animal effects predicted for each genotyped animal. Z-scores were computed standardizing $\hat{a}$ to a mean of zero and variance of one:

$$Z_i = \frac{a_i - \bar{a}}{SD(\hat{a})}$$

where $\bar{a}$ and $SD(\hat{a})$ denote the mean and standard deviation of $\hat{a}$, respectively.

RESULTS

Sequencing, Read Mapping, and Gene Expression

RNA-Seq libraries from hypothalamus, duodenum, ileum, and jejunum tissue of 30 pigs with divergent feed efficiency phenotypes were sequenced, generating over 7.4 billion 75-bp paired-end reads, with an average of 61.8 million reads per library (**Table 2**). After adapter removal and read trimming, the resulting high-quality reads were mapped to the Sscrofa 11.1 genome assembly (NCBI accession AEMK00000000.2) with an average 98.6% read mapping rate per library. Sequencing statistics for individual libraries are given in **Table S1**.

Normalized gene expression values were computed for the 29,651 annotated genes in the porcine genome, and lowly expressed genes across the six tissues were removed, resulting in a set of 19,365 genes to be used in downstream analyses. **Table 3** shows the number of genes expressed in each of the

TABLE 2 | Summary of sequencing statistics by tissue.

Tissue	Total number reads	Mean number reads per library	Mean read mapping % per library
Hypothalamus	2,014,157,388	67,138,579.6	96.91%
Duodenum	1,653,423,084	55,114,102.8	98.25%
Jejunum	1,809,768,258	60,325,608.6	99.59%
Ileum	1,942,804,030	64,760,134.3	99.47%
All	7,420,152,760	61,834,606.3	98.55%

TABLE 3 | Summary of expressed genes by tissue.

Tissue	Total number of genes expressed[1]
Hypothalamus	14,205
Duodenum	12,824
Jejunum	12,734
Ileum	13,640

[1]*Genes defined as expressed if normalized expression ≥ 100 in at least 15 libraries.*

tissues, where a gene is considered expressed if normalized expression ≥ 100 in at least fifteen (half) of the libraries in the tissue. An average of 13,351 genes were expressed per tissue.

Expression Modules Across Individuals and Tissues

A three-way gene x individual x tissue clustering analysis was performed, using constrained tensor decomposition, to obtain a total of 10 gene expression modules.

Module I – Shared, Global Expression

In the first gene expression module, the eigen-tissue and eigen-individual loading distributions are essentially flat (**Figure 1I**). Hence this module captures baseline, global gene expression common to all samples in all tissues. Enrichment analysis showed that many GO terms related to basic eukaryotic cell activities were enriched in the set of 1,307 top genes, including ion binding, protein binding, nucleotide binding, and transport (**Table S2**).

Module II – Hypothalamus

The second gene expression module clearly separated the hypothalamus from the intestinal tissues (**Figure 1II**). In the eigen-individual, more of the proportional variance in loading values was explained by ADFI than contemporary group or gender (6.8% compared to 1.15% and 0.96%, respectively; **Table 4**). The top 130 genes were enriched for functions related to nucleotide binding, protein binding, ion binding, hydrolase activity, and glutamate transporter activity.

Module III – Proximal Small Intestine

The third component captures expression specific to tissues in the proximal small intestine, the duodenum and jejunum. The eigen-tissue is primarily driven by the duodenum (**Figure 1III**). A moderate amount of variation among individuals was explained by both gender (8.13%) and ADFI (5.58%), while the variance explained by contemporary group was negligible (~ 0%). A total of 88 genes passed the thresholding to be considered a top gene in the module. These genes were primarily enriched for binding GO terms, including G protein-coupled receptor binding, sulfur compound binding, carbohydrate derivative binding, bile acid binding, cytoskeletal protein binding, ubiquitin protein ligase binding, nucleotide binding, and metal ion binding. Nearly 83% (73/88) of the top genes were also identified as top genes in the hypothalamus expression module (Module II).

Module IV – Distal Small Intestine (positive loadings)

The fourth gene expression module was comprised of the distal small intestinal tissues, the jejunum and ileum, with the ileum being the main driver (larger loading value; **Figure 1IV**). Although contemporary group explained the largest amount of proportional variance (10.53%), a moderate amount of variation, 6.05%, was explained by ADFI. Top genes in the module were enriched for functions related to peptide transport, lipid transport, chemokine receptor binding, hydrolase activity, bile acid binding, peptidase inhibition, and ion binding. Only one of the top genes, *COX1*, overlapped with the top genes from Module II, while 15 genes from Module III's top set were overlapped.

Module V – Jejunum

Expression in the jejunum tissue was captured in the fifth component (**Figure 1V**). Contemporary group was the only covariate to account for more than 1% of the variation among individuals. GO analysis of the 121 top genes identified that translation regulation, RNA binding, fatty acid binding, and rRNA binding were significantly enriched.

Module VI – Jejunum and Hypothalamus (negative loadings)

The sixth module included the hypothalamus and the jejunum in the eigen-gene, with the jejunum tissue having a much stronger effect (**Figure 1VI**). Again, contemporary group was the main covariate explaining individual loading value variation, as it explained approximately 6% of the variation and ADFI and gender each explained less than 1%. No GO terms were significantly enriched in the set of top genes.

Module VII – Small Intestine

Expression in all three parts of the small intestine, the duodenum, jejunum, and ileum, was captured in the seventh module. The duodenum was the most significant driver, while the jejunum and ileum had very similar loading values (**Figure 1VII**). Once again, variation in loading values in the eigen-individual was predominantly explained by contemporary group.

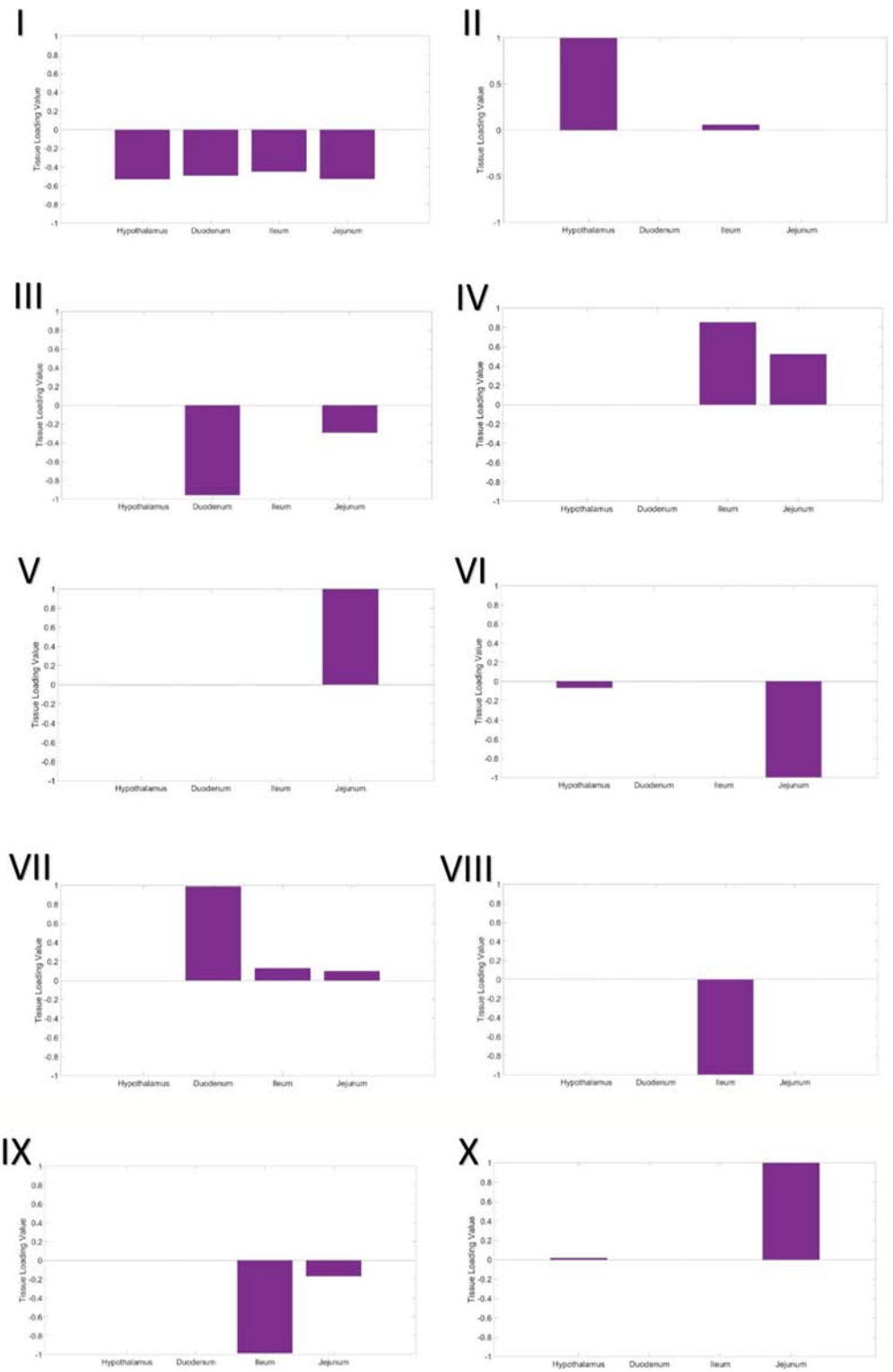


FIGURE 1 | Tissue loading values for Modules I–X from the tensor decomposition.

TABLE 4 | Proportional variance in individual loading values explained by average daily feed intake (ADFI), contemporary group, and gender in each of the modules obtained from the tensor decomposition.

Module	ADFI	Cont. Group	Gender
I	2.02%	3.46%	3.28%
II	6.80%	1.15%	0.96%
III	5.58%	0.00%	8.13%
IV	6.05%	10.53%	0.01%
V	0.00%	10.98%	0.67%
VI	0.36%	6.08%	0.14%
VII	0.00%	18.01%	0.42%
VIII	1.97%	1.47%	2.11%
IX	2.98%	10.09%	0.00%
X	5.05%	21.22%	0.26%

The GO term CMP-N-acetylneuraminate monooxygenase activity was significantly enriched in the top genes.

Module VIII – ileum

Ileum gene expression was highlighted in the eighth component (**Figure 1VIII**). Variation between individual loading values was not well-explained by any of the covariates in the model, ADFI (1.97%), contemporary group (1.47%), and gender (2.11%). No GO terms were significantly enriched in the set of top genes. Additionally, no top genes were overlapped with those from Module IV, which was also driven by gene expression in the ileum.

Module IX – distal small intestine (negative loadings)

The fourth gene expression module was comprised of the distal small intestinal tissues, the jejunum and ileum (**Figure 1IX**). It should be noted that this module corresponds to negative loading values for the tissues, while the results in Module IV corresponded to positive loading values. Similar to Module IV, ileum was the main driver of expression in the module, and contemporary group explained the largest amount of variation between individual loadings. However, none of the top genes were found to be top genes in Module IV, and no GO terms were significantly enriched.

Module X – jejunum and hypothalamus (positive loadings)

The sixth module included the hypothalamus and the jejunum in the eigen-gene, with the jejunum tissue having a much stronger effect (**Figure 1X**). This module corresponds to positive loading values in the eigen-tissue, while Module VI gave the results for negative loading values. A larger amount of variation among individuals was explained by covariates in the model than that from Module IV, contemporary group explained 21.22% and ADFI explained 5.05%. The GO term CMP-N-acetylneuraminate monooxygenase activity was significantly enriched in the top genes. There was no overlap between the set of top genes and the top genes from Module IV.

Genetic Association Analysis

For each of the gene modules, three genetic association analyses were conducted: (1) a weighted analysis with all SNP, (2) an unweighted analysis using only SNP with weight > 5, and (3) a weighted analysis using only SNP with weight > 5. Removal of low weight SNP resulted in SNP sets ranging in size from 101 to 944 markers (**Table 5**). Results from these analyses are shown in **Tables 5** and **6**. Utilization of all 49,691 SNP with pedigree and genomic relationships resulted in heritabilities of 0.366 and 0.269, respectively. In general, applying SNP weights derived from each of the gene models resulted in heritabilities that remained close to those derived from the unweighted pedigree and genomic models (**Table 7**).

The removal of SNP with weight < 5 and leaving SNP unweighted in the model decreased performance in all 10 modules (**Table 5**), i.e., heritabilities were below those of the pedigree and unweighted models. Removal of SNP with weight < 5 and utilizing the SNP weights in the model increased performance from the unweighted case in all ten modules, but

TABLE 5 | Heritability estimates for feed efficiency from unweighted genome-wide association studies (GWAS) utilizing SNP with weight > 5.

Data Set	Heritability (h^2)	Standard Error (SE)	# SNP	h^2/# SNP
Module I	0.069	0.016	183	3.77E-04
Module II	0.062	0.015	204	3.03E-04
Module III	0.061	0.016	145	4.21E-04
Module IV	0.088	0.018	944	9.32E-05
Module V	0.088	0.018	536	1.64E-04
Module VI	0.040	0.012	192	2.08E-04
Module VII	0.045	0.013	101	4.46E-04
Module VIII	0.081	0.017	528	1.53E-04
Module IX	0.081	0.017	624	1.30E-04
Module X	0.015	0.014	296	5.07E-05

TABLE 6 | Heritability estimates for feed efficiency from weighted genome-wide association studies (GWAS) utilizing SNP with weight > 5.

Data Set[1]	Heritability (h^2)[2]	Standard Error (SE)	# SNP	h^2/# SNP
Module I	0.106	0.049	183	5.81E-04
Module II	0.099	0.050	204	4.86E-04
Module III	0.094	0.053	145	6.52E-04
Module IV	0.137	0.0374	944	1.45E-04
Module V	0.156	0.026	536	2.91E-04
Module VI	0.073	0.072	192	3.82E-04
Module VII	0.081	0.066	101	7.98E-04
Module VIII	0.131	0.036	528	2.49E-04
Module IX	0.128	0.040	624	2.04E-04
Module X	0.089	0.061	296	3.00E-04

[1]*SNP weights derived from indicated gene module.*
[2]*Heritability estimates were corrected using the difference between the phenotypic variance estimated with the unweighted G and residual variance estimated with each weighted G.*

TABLE 7 | Heritability estimates for feed efficiency from weighted genome-wide association studies (GWAS) utilizing all SNP.

Data Set[1]	Heritability (h^2)[2]	Standard Error (SE)	# SNP	h^2/# SNP
Pedigree (Unweighted)	0.366	0.045	49,691	7.37E-06
Genomic (Unweighted)	0.269	0.031	49,691	5.42E-06
Module I	0.270	0.035	49,691	5.44E-06
Module II	0.271	0.037	49,691	5.46E-06
Module III	0.269	0.028	49,691	5.41E-06
Module IV	0.273	0.040	49,691	5.49E-06
Module V	0.272	0.038	49,691	5.46E-06
Module VI	0.270	0.034	49,691	5.43E-06
Module VII	0.269	0.026	49,691	5.40E-06
Module VIII	0.272	0.039	49,691	5.49E-06
Module IX	0.273	0.039	49,691	5.49E-06
Module X	0.270	0.035	49,691	5.44E-06

[1]*SNP weights derived from indicated gene module.*
[2]*Heritability estimates were corrected using the difference between the phenotypic variance estimated with the unweighted G and residual variance estimated with each weighted G.*

overall heritability was still lower than that obtained from using all SNP (**Table 6**).

Output from the association analyses for feed intake is shown in **Table S3**. A total of 36 unique significant SNP associations were identified across the ten gene modules, while 2 only SNP were significant in the standard analysis using all 49,691 SNP with no SNP weights. Neither of the 2 SNP identified in the unweighted analysis were identified in the weighted analyses. The number of significant SNP identified in each module's analysis ranged from 0 to 22, with Modules I, II, VI, VII, and X having only no significant SNP and Module III having 22 significant SNP. For the weighted analyses, significant SNP were identified on chromosomes SSC 2, 4, 5, 7, 8, 9, 13, 14, 15, 18, and X, with SSC 9 and SSC 8 having the largest numbers of significant SNP, 12 and 6, respectively.

DISCUSSION

The most widely used approach to GWAS has been to assign equal prior probability of association to all sequence variants tested. Recent findings suggest that incorporating prior information can increase the power for identifying associations. Such prior information can be obtained from several different sources, including but not limited to linkage analysis (Roeder et al., 2006), gene expression (Li et al., 2013; Gamazon et al., 2015; Gusev et al., 2016; Xu et al., 2017), and functional annotation of variants (Sveinbjornsson et al., 2016). In this work, we present a methodology that exploits multi-tissue transcriptional data from a small set of individuals with extreme phenotypes to assign SNP weights for a GWAS on an expanded set of phenotyped individuals. It has been shown that any set of nonnegative weights can guarantee substantial power gain if the weights are informative and little power loss if the weights are uninformative (Genovese et al., 2006). Hence, the weighting procedure is robust to the informativeness of the weights.

We applied our method to identify genetic markers associated with feed intake in swine. The gut-brain axis is comprised of bidirectional communication between the central and enteric nervous systems, linking cognitive centers of the brain with peripheral intestinal functions. The gut-brain axis modulates short-term satiety and hunger responses to regulate the delivery of nutrients and transit of nutrients through the gastrointestinal tract (Hussain and Bloom, 2012). RNA-Seq was performed on tissues involved in the gut-brain axis, including hypothalamus, duodenum, ileum, and jejunum, originating from pigs with extreme feed intake phenotypes. A tensor decomposition method, which performs three-way clustering across genes,

tissues, and individuals, was used to identify gene expression modules that were either common to all tissues and individuals or exclusive to particular tissue/individual combinations.

The top ten gene modules from the tensor decomposition were considered. Note that since the clustering algorithm generates expression modules *via* successive rank-1 approximations, if more expression modules were desired the algorithm could simply be applied to the residual tensor. Module I captured baseline, global gene expression common to all samples in all tissues, indicated by the flat distributions of the eigen-tissue and eigen-individual loading values. Other gene modules captured expression specific portions of the gut-brain axis, including the hypothalamus, the proximal and distal small intestine, the entire small intestine, and the individual components of the small intestine.

A tensor projection model was used to identify ADFI-associated genes within each of the ten modules. The P-values obtained from testing the ADFI-effect were used to weight SNP in order to conduct a weighted GWAS. P-values were chosen over regression coefficients for weighting in order to rank SNP according to the significance of their respective genomic regions rather than simply an effect size. Results from both the weighted and unweighted analyses are shown in **Tables 7**–**9**. Preliminary analyses using weighted SNP revealed what appeared to be inflated estimates of heritability. There was substantially less change in residual variance estimates, indicating that inflated heritability was not a result of explaining substantially more phenotypic variation with the weighted G, but an artifact of weighted G resulting in inflated additive and phenotypic variance estimates. Phenotypic variance should remain constant, so heritability estimates were corrected using the difference between the phenotypic variance estimated with the unweighted G and residual variance estimated with each weighted G.

There was a common pattern to the change in heritability estimates as the SNP prioritization changed. When using all 50K unweighted SNP, the heritability increased from 0.269 using genomic relationships to 0.366 using pedigree. In all ten modules, the use of weighted SNP restricted to those with weight > 2 resulted in a heritability slightly lower, but comparable to that from the usual unweighted, genomic model. Randomization of SNP weights (**Table S4**) resulted in nearly the same overall and average per SNP heritabilities, suggesting that the weighting threshold may be suboptimal.

To investigate if a more stringent SNP weight threshold could increase model performance, SNP with weight < 5 were removed from the analysis. This resulted in an average 21-fold drop in the number of SNP included in each analysis (**Table 5**). Although overall heritability estimates were lower than those obtained using SNP with weight > 2, the heritability per SNP increased. Additionally, in most modules, both overall and per SNP heritabilities were higher than those obtained when the SNP weights were randomized. The numbers of SNP ($101 < p < 944$) in these analyses were smaller than the number of animals (n =

TABLE 8 | Heritability estimates for feed efficiency from unweighted genome-wide association studies (GWAS) utilizing SNP with weight > 2.

Data Set	Heritability (h^2)	Standard Error (SE)	# SNP	h^2/# SNP
Module I	0.209	0.028	5,915	3.53E-05
Module II	0.182	0.026	5,259	3.35E-05
Module III	0.139	0.022	1,992	6.98E-05
Module IV	0.221	0.027	7,290	3.03E-05
Module V	0.184	0.025	4,442	4.14E-05
Module VI	0.175	0.025	4,311	4.06E-05
Module VII	0.148	0.022	1,951	7.59E-05
Module VIII	0.229	0.028	9,283	2.47E-05
Module IX	0.225	0.028	7,932	2.84E-05
Module X	0.183	0.025	4,250	4.31E-05

TABLE 9 | Heritability estimates for feed efficiency from weighted genome-wide association studies (GWAS) utilizing SNP with weight > 2.

Data Set[1]	Heritability (h^2)[2]	Standard Error (SE)	# SNP	h^2/# SNP
Module I	0.217	0.036	5,915	3.66E-05
Module II	0.205	0.035	5,259	3.90E-05
Module III	0.166	0.043	1,992	8.31E-05
Module IV	0.255	0.026	7,290	3.50E-05
Module V	0.222	0.028	4,442	5.00E-05
Module VI	0.200	0.037	4,311	4.65E-05
Module VII	0.174	0.041	1,951	8.93E-05
Module VIII	0.250	0.030	9,283	2.69E-05
Module IX	0.238	0.031	7,932	3.00E-05
Module X	0.208	0.034	4,250	4.90E-05

[1]*SNP weights derived from indicated gene module.*

[2]*Heritability estimates were corrected using the difference between the phenotypic variance estimated with the unweighted G and residual variance estimated with each weighted G.*

4,200), eliminating the 'p greater than n' problem. Hence, applying a more stringent threshold results in a more informative set of SNP. Note the weight threshold values of 2 and 5 were chosen arbitrarily. Additional investigation will be needed to determine the optimal weight threshold for SNP inclusion, but this was outside the scope of this study.

Across the ten gene modules (weight > 5), 36 unique SNP were identified as having significant effects, while only 2 SNP were significant in the unweighted analysis utilizing all 50K SNP. Neither SNP from the unweighted analysis resided in known QTL related to swine feed efficiency (feed intake, average daily gain, and feed conversion ratio) compared to 29 (80.6%) in the weighted analyses, with 9 SNP being located in feed intake QTL (**Table S4**). Additionally, many of the genes harboring significant SNP have been identified in previous studies as candidate genes related to feed efficiency in several species (**Table S5**). In particular, the genes *ROBO2* (2 SNP), PLA2G4A (4 SNP), and *MEGF10* (1 SNP) were previously identified as candidate genes for residual feed intake and feed conversion ratio in swine (Ding et al., 2018; Horodyska et al., 2019). Hence, the results from this study suggest that a considerable proportion of heritability of feed intake is driven by many SNP that individually do not attain genome-wide significance in GWAS and therefore support a highly polygenic architecture for feed intake.

Our integrated methodology, at present, is obviously partial to genotyped SNP within genes. Because most available biological resources are biased toward genes, SNP pertaining to known genes likely have more relevant prior information. Consequently, the resulting weights may be more effective for associated SNP residing in or close to known genes. Therefore, results derived from our method can still be informative regardless of their intrinsic bias. Future work will focus on extending the scope of the tensor decomposition step to leverage data from other genomic sources, including but not limited to expression of non-coding RNA, miRNA expression, transcription factors, methylation targets, and miRNA binding. Additionally the method will be extended to prioritize variants from whole genome sequencing for assay development based on functional effects.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI SRA Accession Numbers PRJNA528599, PRJNA528884, PRJNA529214, and PRJNA529662.

AUTHOR CONTRIBUTIONS

BK conceived of the study, and all authors participated in its design and coordination. WO and AP were involved in the acquisition of data, and BK, WS, LK, and GR performed data analyses. BK drafted the manuscript, and AP, WO, LK, WS, and GR contributed to the writing and editing. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Kathy Rohren, Kris Simmerman, and Linda Flathman for technical assistance; and the USMARC Core Laboratory for sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01339/full#supplementary-material

TABLE S1 | Sequencing statistics.

TABLE S2 | Results from Gene Ontology (GO) analysis in each gene module.

TABLE S3 | Significant SNP in association analysis and QTL and genes overlapped.

TABLE S4 | Results of GWAS for SNP weight randomization.

TABLE S5 | Associations identified in previous literature for significant SNP.

REFERENCES

AVMA. (2013). *Guidelines for the Euthanasia of Animals. 2013 ed.* (Schaumburg, IL: AVMA).

Ballester, M., Ramayo-Caldas, Y., Revilla, M., Corominas, J., Castelló, A., Estellé, J., et al. (2017). Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs. *Sci. Rep.* 7, 46539. doi: 10.1038/srep46539

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cai, Z., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2018). Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics* 19, 656. doi: 10.1186/s12864-018-5050-x

Deng, T., Liang, A., Liang, S., Ma, X., Lu, X., Duan, A., et al. (2019). Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Front. Genet.* 10, 36. doi: 10.3389/fgene.2019.00036

Diao, G., and Vidyashankar, A. N. (2013). Assessing genome-wide statistical significance for large *p* small *n* problems. *Genetics* 192, 781–783. doi: 10.1534/genetics.113.150896

Ding, R., Yang, M., Wang, X., Quan, J., Zhuang, Z., Zhou, S., et al. (2018). Genetic architecture of feeding behavior and feed efficiency in a Duroc pig population. *Front. Genet.* 9, 220. doi: 10.3389/fgene.2018.00220

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nuc. Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207

Fang, L., Sahana, G., Su, G., Yu, Y., Zhang, S., Lund, M. S., et al. (2017). Integrating sequence-based GWAS and RNA-Seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. *Sci. Rep.* 7, 45560. doi: 10.1038/srep45560

Fernando, R., Toosi, A., Wolc, A., Garrick, D., and Dekkers, J. (2017). Application of whole-genome prediction methods for genome-wide association studies: a Bayesian approach. *JABES* 22 (2), 172–193. doi: 10.007/s13253-017-0277-6

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). A gene-based association method for mapping traits

using reference transcriptome data. *Nat. Genet.* 47, 1091–1098. doi: 10.1038/ng.3367

Genovese, C. R., Roeder, K., and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* 93 (3), 509–524. doi: 10.1093/biomet/93.3.509

Gusev, A., Ko, A., Shi., H., Bhatia, G., Chung, W., Penninx, B. J. W. H., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi: 10.1038/ng.3506

Horodyska, J., Reyer, H., Wimmers, K., Trakooljul, N., Lawlor, P. G., and Hamill, R. M. (2019). Transcriptome analysis of adipose tissue from pigs divergent in feed efficiency reveals alteration in gene networks related to adipose growth, lipid metabolism, extracellular matrix, and immune response. *Mol. Genet. Genomics* 294, 395–408. doi: 10.1007/s00438-018-1515-5

Hussain, S. S., and Bloom, S. R. (2012). The regulation of food intake by the gut-brain axis: implications for obesity. *Int. J. Obes.* 37, 625–633. doi: 10.1038/ijo.2012.93

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kommadath, A., Bao, H., Choi, I., Reecy, J. M., Koltes, J. E., Fritz-Waters, E., et al. (2017). Genetic architecture of gene expression underlying variation in host response to porcine reproductive and respiratory syndrome virus infection. *Sci. Rep.* 7, 46203. doi: 10.1038/srep46203

Li, L., Kabesch, M., Bouzigon, E., Demenais, F., Farrall, M., Moffatt, M. F., et al. (2013). Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. *Front. Genet.* 4, 103. doi: 10.3389/fgene.2013.00103

Love, M., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1. doi: 10.1186/s13059-014-0550-8

Meyer, K. (2007). WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. *J. Zhejiang Uni. Sci. B* 8, 815–821. doi: 10.1631/jzus.2007.B0815

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 44 (D1), D336–D342. doi: 10.1093/nar/gkv1194

NRC (2012). *Nutrient Requirements of Swine. 11th ed.* (Washington, DC: National Academy Press).

Oliver, W. T., Mathews, S. A., Phillips, O., Jones, E. E., Odle, J., and Harrell, R. J. (2002). Efficacy of partially hydrolyzed corn syrup solids as a replacement for lactose in manufactured liquid diets for neonatal pigs. *J. Anim. Sci.* 80, 143–153. doi: 10.2527/2002.801143x

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome for RNA-seq reads. *Nat. Biotech.* 33, 290–295. doi: 10.1038/nbt.3122

Roeder, K., Bacanu, S. A., Wasserman, L., and Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* 78, 243–252. doi: 10.1086/500026

Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S. A., Oddson, A., Másson, G., et al. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* 48, 314–317. doi: 10.1038/ng.3507

Thorson, J. F., Heidorn, N. L., Ryu, V., Czaja, K., Nonneman, D. J., Barb, C. R., et al. (2017). Relationship of neuropeptide FF receptors with pubertal maturation of gilts. *Biol. Reprod.* 96, 617–634. doi: 10.1095/biolreprod.116.144998

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., et al. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* 96 (1), 668–678. doi: 10.3168/jds.2012-5702

VanRaden, P. M. (2008). Efficient methods to comput genomic predictions. *J. Dairy Sci.* 91 (11), 4414–4423. doi: 10.3168/jds.2007-0980

Wang, H., Misztal, I., Agiular, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)* 94 (2), 73–83. doi: 10.1017/S0016672312000274

Wang, M., Fischer, J., and Song, Y. S. (2017). Three-way clustering of multi-tissue multi-individual gene expression data using constrained tensor decomposition. *bioRxiv*. doi: 10.1101/229245

Xu, Z., Wu, C., Wei, P., and Pan, W. (2017). A powerful framework for integrating eQTL and GWAS summary data. *Genetics* 207 (3), 893–902. doi: 10.1534/genetics.117.300270

Disclaimer: Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Keel, Snelling, Lindholm-Perry, Oliver, Kuehn and Rohrer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 February 2020
doi: 10.3389/fgene.2019.01386

Alveolar Macrophage Chromatin Is Modified to Orchestrate Host Response to *Mycobacterium bovis* Infection

Thomas J. Hall[1], Douglas Vernimmen[2], John A. Browne[1], Michael P. Mullen[3], Stephen V. Gordon[4,5], David E. MacHugh[1,5]* and Alan M. O'Doherty[1]*

[1] Animal Genomics Laboratory, UCD School of Agriculture and Food Science, College Dublin, Dublin, Ireland, [2] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom, [3] Bioscience Research Institute, Athlone Institute of Technology, Athlone, Ireland, [4] UCD School of Veterinary Medicine, University College Dublin, Dublin, Ireland, [5] UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:
Jiuzhou Song,
University of Maryland,
College Park, United States

Reviewed by:
Luyang Sun,
Baylor College of Medicine,
United States
Ying Yu,
China Agricultural University
(CAU), China

***Correspondence:**
Douglas Vernimmen
douglas.vernimmen@roslin.ed.ac.uk
David E. MacHugh
david.machugh@ucd.ie

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 06 June 2019
Accepted: 18 December 2019
Published: 07 February 2020

Citation:
Hall TJ, Vernimmen D, Browne JA, Mullen MP, Gordon SV, MacHugh DE and O'Doherty AM (2020) Alveolar Macrophage Chromatin Is Modified to Orchestrate Host Response to Mycobacterium bovis Infection. Front. Genet. 10:1386. doi: 10.3389/fgene.2019.01386

Bovine tuberculosis is caused by infection with *Mycobacterium bovis*, which can also cause disease in a range of other mammals, including humans. Alveolar macrophages are the key immune effector cells that first encounter *M. bovis* and how the macrophage epigenome responds to mycobacterial pathogens is currently not well understood. Here, we have used chromatin immunoprecipitation sequencing (ChIP-seq), RNA-seq and miRNA-seq to examine the effect of *M. bovis* infection on the bovine alveolar macrophage (bAM) epigenome. We show that H3K4me3 is more prevalent, at a genome-wide level, in chromatin from *M. bovis*-infected bAM compared to control non-infected bAM; this was particularly evident at the transcriptional start sites of genes that determine programmed macrophage responses to mycobacterial infection (e.g. M1/M2 macrophage polarisation). This pattern was also supported by the distribution of RNA Polymerase II (Pol II) ChIP-seq results, which highlighted significantly increased transcriptional activity at genes demarcated by permissive chromatin. Identification of these genes enabled integration of high-density genome-wide association study (GWAS) data, which revealed genomic regions associated with resilience to infection with *M. bovis* in cattle. Through integration of these data, we show that bAM transcriptional reprogramming occurs through differential distribution of H3K4me3 and Pol II at key immune genes. Furthermore, this subset of genes can be used to prioritise genomic variants from a relevant GWAS data set.

Keywords: ChIP-seq, chromatin, integrative genomics, macrophage, microRNA-seq, *Mycobacterium bovis*, RNA-seq, tuberculosis

Abbreviations: bAM: bovine alveolar macrophage/s; bTB: bovine tuberculosis; ChIP: chromatin immunoprecipitation; FDR: false discovery rate; GWAS: genome-wide association study; H3K27me3: histone H3 lysine 27 tri-methylation; H3K4me3: histone H3 lysine 4 tri-methylation; hpi: hours post infection; $\log_2$FC: $\log_2$ fold change; Pol II: RNA Polymerase II; SNP: single-nucleotide polymorphism; TB: tuberculosis.

INTRODUCTION

Bovine tuberculosis (bTB) is a chronic infectious disease of livestock, particularly domestic cattle (*Bos taurus*, *Bos indicus* and *Bos taurus*/*indicus* hybrids), which causes more than $3 billion in losses to global agriculture annually (Steele, 1995; Waters et al., 2012). The aetiological agent of bTB is *Mycobacterium bovis*, a pathogen with a genome sequence that is 99.95% identical to *M. tuberculosis*, the primary cause of human tuberculosis (TB) (Garnier et al., 2003). In certain agroecological milieus *M. bovis* can also cause zoonotic TB with serious implications for human health (Thoen et al., 2016; Olea-Popelka et al., 2017; Vayr et al., 2018).

Previous studies have shown that the pathogenesis of bTB disease in animals is similar to TB disease in humans and many of the features of *M. tuberculosis* infection are also characteristic of *M. bovis* infection in cattle (Waters et al., 2014; Buddle et al., 2016; Williams and Orme, 2016). Transmission is *via* inhalation of contaminated aerosol droplets and the primary site of infection is the lungs where the bacilli are phagocytosed by alveolar macrophages, which normally can contain or destroy intracellular bacilli (Weiss and Schaible, 2015; Kaufmann and Dorhoi, 2016). Disease-causing mycobacteria, however, can persist and replicate within alveolar macrophages *via* a bewildering range of evolved mechanisms that subvert and interfere with host immune responses (de Chastellier, 2009; Cambier et al., 2014; Schorey and Schlesinger, 2016; Awuh and Flo, 2017). These mechanisms include recruitment of cell surface receptors on the host macrophage; blocking of macrophage phagosome–lysosome fusion; detoxification of reactive oxygen and nitrogen intermediates (ROI and RNI); harnessing of intracellular nutrient supply and metabolism; inhibition of apoptosis and autophagy; suppression of antigen presentation; modulation of macrophage signalling pathways; cytosolic escape from the phagosome; and induction of necrosis, which leads to immunopathology and shedding of the pathogen from the host (Ehrt and Schnappinger, 2009; Hussain Bhat and Mukhopadhyay, 2015; Queval et al., 2017; BoseDasgupta and Pieters, 2018; Chaurasiya, 2018; Stutz et al., 2018).

Considering the dramatic perturbation of the macrophage by intracellular mycobacteria, we and others have demonstrated that bovine and human alveolar macrophage transcriptomes are extensively reprogrammed in response to infection with *M. bovis* and *M. tuberculosis* (Nalpas et al., 2015; Vegh et al., 2015; Lavalett et al., 2017; Jensen et al., 2018; Malone et al., 2018; Papp et al., 2018). These studies have also revealed that differentially expressed gene sets and dysregulated cellular networks and pathways are functionally associated with many of the macrophage processes described above that can control or eliminate intracellular microbes.

For many intracellular pathogens, it is now also evident that the infection process involves alteration of epigenetic marks and chromatin remodelling that may profoundly alter host cell gene expression (Hamon and Cossart, 2008; Bierne et al., 2012; Rolando et al., 2015; Niller and Minarovits, 2016). For example, distinct DNA methylation changes are detectable in macrophages infected with the intracellular protozoan *Leishmania donovani*, which causes visceral leishmaniasis (Marr et al., 2014). Recent studies using cells with a macrophage phenotype generated from the THP-1 human monocyte cell line have provided evidence that infection with *M. tuberculosis* induces alterations to DNA methylation patterns at specific inflammatory genes (Zheng et al., 2016) and across the genome in a non-canonical fashion (Sharma et al., 2016).

With regards to host cell histones and in the context of mycobacterial infections, Yaseen et al. (2015) have shown that the Rv1988 protein, secreted by virulent mycobacteria, localises to the chromatin upon infection and mediates repression of host cell genes through methylation of histone H3 at a non-canonical arginine residue. In addition, chromatin immunoprecipitation sequencing (ChIP-seq) analysis of H3K4 monomethylation (a marker of poised or active enhancers) showed that regulatory sequence motifs embedded in subtypes of Alu SINE transposable elements are key components of the epigenetic machinery modulating human macrophage gene expression during *M. tuberculosis* infection (Bouttier et al., 2016).

In light of the profound macrophage reprogramming induced by mycobacterial infection, and previous work demonstrating a role for host cell chromatin modifications, we have used ChIP-seq and RNA sequencing (RNA-seq) to examine gene expression changes that reflect host–pathogen interaction in bovine alveolar macrophages (bAM) infected with *M. bovis*. The results obtained support an important role for dynamic chromatin remodelling in the macrophage response to mycobacterial infection, particularly with respect to M1/M2 polarisation. Genes identified from ChIP-seq and RNA-seq results were also integrated with genome-wide association study (GWAS) data to prioritise genomic regions and single-nucleotide polymorphisms (SNPs) associated with bTB resilience. Finally, the suitability of bAM for ChIP-seq assays and the results obtained demonstrate that these cells represent an excellent model system for unravelling the epigenetic and transcriptional circuitry perturbed during mycobacterial infection of vertebrate macrophages.

MATERIALS AND METHODS

Preparation and Infection of bAM

bAM and *M. bovis* 2122 were prepared as described previously (Magee et al., 2014) with minor adjustments. Macrophages (2×10^6) were seeded in 60 mm tissue culture plates and challenged with *M. bovis* at a multiplicity of infection (MOI) of 10:1 (2×10^7 bacteria per plate) for 24 h; parallel non-infected controls were prepared simultaneously.

Preparation of Nucleic Acids for Sequencing

Sheared fixed chromatin was prepared exactly as described in the truChIP™ Chromatin Shearing Kit (Covaris) using 2×10^6 macrophage cells per AFA tube. Briefly, cells were washed in cold PBS and 2.0 ml of Fixing Buffer A was added to each plate, to which 200 µl of freshly prepared 11.1% formaldehyde solution was added. After 10 min on a gentle rocker the crosslinking was halted by the addition of 120 µl of Quenching Solution E; cells

were washed with cold PBS, released from the plate using a cell scraper and resuspended in 300 µl Lysis Buffer B for 10 min with gentle agitation at 4°C to release the nuclei. The nuclei were pelleted and washed once in Wash Buffer C and three times in Shearing Buffer D3 prior to being resuspended in a final volume of 130 µl of Shearing Buffer D3. The nuclei were transferred to a micro AFA tube and sonicated for 8 min each using the Covaris E220e as per the manufacturer's instructions. Chromatin immunoprecipitation of sonicated DNA samples was carried out using the Chromatin Immunoprecipitation (ChIP) Assay Kit (Merck KGaA) and anti-H3K4me3 (05-745R) (Merck KGaA), Pol II (H-224) (Santa Cruz Biotechnology, Inc.) or anti-H3K27me3 (07-449) (Merck KGaA) as previously described (Vernimmen et al., 2011). RNA was extracted from infected ($n = 4$) and control ($n = 4$) bAM samples using the RNeasy Plus Mini Kit (Qiagen) as previously described (O'Doherty et al., 2012). All eight samples exhibited excellent RNA quality metrics (RIN >9).

Sequencing

Illumina TruSeq Stranded mRNA and TruSeq Small RNA kits were used for mRNA-seq and small RNA-seq library preparations and the NEB Next Ultra ChIPseq Library Prep kit (New England Biolabs) was used for ChIP-seq library preparations. Pooled libraries were sequenced by Edinburgh Genomics (http://genomics.ed.ac.uk) as follows: paired-end reads (2 × 75 bp) were obtained for mRNA and ChIP DNA libraries using the HiSeq 4000 sequencing platform and single-end read (50 bp) were obtained for small RNA libraries using the HiSeq 2500 high output version 4 platform.

ChIP-seq Bioinformatics Analysis

Computational analyses for all bioinformatic processes were performed on a 72-CPU computer server with Linux Ubuntu (version 16.04.4 LTS). An average of 54 M paired end 75 bp reads were obtained for each histone mark. At each step of data processing, read quality was assessed *via* FastQC (version 0.11.5) (Andrews, 2016). Any samples that indicated adapter contamination were trimmed *via* Cutadapt (version 1.15) (Martin, 2011). Correlation plots generated with EaSeq (version 1.05) (Lerdrup et al., 2016) of genome-wide H3K4me3, H3K27me3 and Pol II sequencing reads from infected and non-infected bAM showed high correlation between samples (Pearson's correlation coefficient: 0.93–0.97) for all three ChIP-seq targets (**Supplementary Figure 1**). After data quality control and filtering, ~760 million paired end reads were aligned to the UMD 3.1 bovine genome assembly using Bowtie2 (version 2.3.0) (Langmead and Salzberg, 2012). The mean alignment rate for the histone marks was 96.23%. The resulting SAM files were converted and indexed into BAM files *via* Samtools (version 1.3.1) (Li et al., 2009). After alignment, samples were combined and sorted into 14 files, based on the animal (A1 or A2), the histone mark (K4/K27/Pol II) and treatment (control or infected), i.e. A1-CTRL-K4. Peaks were called by using alignment files to determine where the reads have aligned to specific regions of the genome, and then comparing that alignment to the input samples as a normalisation step.

The peak calling was carried out *via* MACS2 (version 2.1.1.20160309) (Feng et al., 2011). The H3K4me3 mark was called in sharp peak mode and H3K27me3 and Pol II were called in broad peak mode, as per the user guide. Peak tracks were generated in MACS2 and visualised with the Integrative Genome Viewer (version 2.3) (Thorvaldsdottir et al., 2013). Union peaks were generated by combining and merging overlapping peaks in all samples for each histone mark. Differential peak calling was called *via* MACS2 using the bdgdiff function. Peak images were generated by visually assessing all three marks in tandem across the entire bovine genome with the Integrative Genomics Viewer (IGV). The significance of peaks was determined by sorting peaks for each mark in each treatment by *P* value and then fold enrichment with a cut-off of 2.0 and a *P* value threshold of 0.05 (Wilbanks and Facciotti, 2010). Peaks from each animal in each condition for each mark were cross-referenced with the IGV images and differential peak caller to determine a difference in fold enrichment for each observed peak difference between conditions. This required comparing peak start and end sites, chromosomes, *P* and *q* values for each summit, summit locations and normalised fold enrichment of a peak against the input sample (see **Supplementary Information File 1** for peak sets). Any peaks that exhibited a difference of 4 or greater fold enrichment, a *P* value of less than 0.05, a false discovery rate (FDR) less than 0.05 and that were also identified by the differential peak caller were selected for further analysis [see **Supplementary Information File 1** for peaks at transcription start sites (TSSs) that met some but not all of the above criteria]. Peaks that were then classified to be different between conditions in all three data sets were examined to determine their proximity to TSS. Differential peaks were also called using the R package DiffBind (version 2.80) (Stark and Brown, 2011). DiffBind includes functions to support the processing of peak sets, including overlapping and merging peak sets, counting sequencing reads overlapping intervals in peak sets and identifying statistically significantly differentially bound sites based on evidence of binding affinity (measured by differences in read densities; see **Supplementary Information File 1**). For H3K27me3 DiffBind differential peak calling, the initial MACS2 peak list, consisting of 64,264 total peaks (see **Supplementary Information File 1**), was merged and reduced to a smaller group of larger, broader peaks to reduce noise and false positive discovery (**Figure 2B**).

RNA-Seq Bioinformatics Analysis

An average of 44 M paired end 75 bp reads were obtained for each of the eight samples (four control, four infected). Adapter sequence contamination and paired-end reads of poor quality were removed from the raw data. At each step, read quality was assessed with FastQC (version 0.11.5). Any samples that indicated adapter contamination were trimmed *via* Cutadapt (version 1.15). After quality control and filtering, ~250 million reads were mapped to the bovine genome, with 72% total read mapping, overall. The raw reads were aligned to the UMD 3.1.1 bovine transcriptome using Salmon (version 0.8.1) (Patro et al., 2017). Aligned reads were also counted in Salmon and the resulting quantification files were annotated at gene level *via* tximport (version 3.7) (Soneson et al., 2015). The annotated gene

counts were then normalised and differential expression analysis performed with DESeq2 (version 1.20.0) (Love et al., 2014), correcting for multiple testing using the Benjamini–Hochberg method (Benjamini and Hochberg, 1995). Genes identified from ChIP-seq as exhibiting differential histone modifications were cross-referenced with the RNA-seq data set to determine significant $\log_2$FC between *M. bovis*-infected and control non-infected. Additionally, this RNA-seq data was cross-referenced with RNA-seq data from a previous study that investigated bAM infected with *M. bovis* (Nalpas et al., 2015).

MicroRNA-Seq Bioinformatics Analysis

A mean of 26 M paired-end 50 bp reads were obtained for each of the eight samples (four control, four infected). At each step of data processing, read quality was assessed *via* FastQC (version 0.11.5). Any samples that exhibited adapter contamination were trimmed *via* Cutadapt (version 1.15) and all reads smaller than 17 bp were removed from the analysis. After quality control and filtering, ~100 million reads were mapped to the bovine genome, with 79% total reads mapping, overall. Raw reads were mapped to UMD3.1 using Bowtie (version 1.2.2). miRNA detection, identification and quantification were carried out with mirdeep2 (version 0.0.91). Isoform analysis was also performed using mirdeep2. Differential expression analysis was performed using DESeq2, correcting for multiple testing with the Benjamini–Hochberg method. Any miRNAs that were significantly differentially expressed (FDR < 0.10) were selected for further analysis. To determine if significantly differentially expressed miRNAs target genes selected in the ChIP-seq analysis, miRmap (Vejnar and Zdobnov, 2012) was used to predict the likelihood that a specific miRNA targets one or more of the genes based on three criteria: delta G binding, probability exact and phylogenetic conservation of seed site, which is then combined into a single scoring metric (miRmap score). Any predicted gene targets with miRmap score ≥0.70 were included in the analysis (see **Supplementary Information File 3**).

Pathway Analysis

Pathway analysis was carried out on any gene that had a differential peak between control and infected samples. Pathway analysis and gene ontology (GO) summarisation was carried out using DAVID (version 6.8), Ingenuity Pathway Analysis—IPA (version 1.1, winter 2018 release) and PANTHER (version 13.1) (Kramer et al., 2014; Mi et al., 2017). KEGG pathways were selected by choosing pathways that had the highest number of genes identified in the ChIP-seq data and had an FDR < 0.05.

Integration of GWAS Data

GWAS data for genetic susceptibility to *M. bovis* infection previously generated by Richardson et al. (2016) were analysed to determine if subsets of SNPs selected according to their distance to H3K4me3 and Pol II active loci were enriched for significant GWAS hits. The nominal *P* values used in this study were generated using single SNP regression analysis in a mixed animal model as described previously (Richardson et al., 2016). In summary, high-density genotypes (n = 597,144) of dairy bulls (n = 841) used for artificial insemination were associated with deregressed estimated breeding values for bTB susceptibility that had been calculated from epidemiological information on 105,914 daughters and provided by the Irish Cattle Breeding Federation (ICBF). In this study, the significance of the distribution of SNP nominal *P* values (from Richardson et al., 2016) within and up to 100 kb up- and downstream to genes identified as having differential H3K4me3 and Pol II activity on bTB susceptibility was estimated in R using *q* value (FDRTOOL) and permutation analysis (custom scripts). A total of 1,000 samplings (with replacement) from the HD GWAS *P* value data set (n = 597,144) representing the size of each of selected SNP subsets were generated. The *q* values for each SNP *P* value subset and all its permuted equivalents were calculated using the FDRTOOL library in R. The subsequent significance level (P_{perm}) assigned to each of the SNP subsets was equivalent to the proportion of permutations in which at least the same number of *q* values < 0.05 as the SNP subset were obtained, i.e. by chance.

RESULTS

M. bovis Infection Induces Trimethylation of H3K4 at Key Immune Function Related Loci in Bovine Alveolar Macrophages

Previous studies have shown that bAM undergo extensive gene expression reprogramming following infection of *M. bovis* (Nalpas et al., 2015; Malone et al., 2018), with almost one half of the detectable transcriptome exhibiting significant differential expression within bovine macrophages 24 h after infection (Nalpas et al., 2015). Changes of this magnitude are comparable to those observed in previous experiments that have examined the chromatin remodelling that accompanies mycobacterial infection of macrophages, where trimethylation of lysine 4 of Histone H3 (H3K4me3) was shown to correlate with active transcription (Bouttier et al., 2016; Arts et al., 2018).

We used ChIP-seq to examine histone modification changes that occur after *M. bovis* infection of bAM from sex- and aged-matched Holstein-Friesian cattle. The aim was to determine genome-wide changes in the distribution of H3K4me3 and H3K27me3, and Pol II occupancy at the response genes (Sims et al., 2003). Differential peaks between conditions were called, compared and visualised with IGV to determine where differences in H3K4me3, H3K27me3 and Pol II occupancy occur between control and infected bAM (**Figure 1**). ChIP-seq peaks are defined as areas of the genome enriched by read counts after alignment to the reference genome.

Peak differences for H3K4me3 occurred at multiple locations across the genome and were estimated by the fold enrichment of a peak normalised against input control DNA that had not undergone antibody enrichment. Differential peaks in each condition were defined by several criteria: 1) the fold enrichment of each peak had to be larger than 10 in at least one condition (Landt et al., 2012); 2) the identified peaks had a *P*-value cut off of 0.05; 3) the peaks being compared in each condition were no more than 500 bp up- and downstream of each other; 4) the peaks were classified as different using log-likelihood ratios and

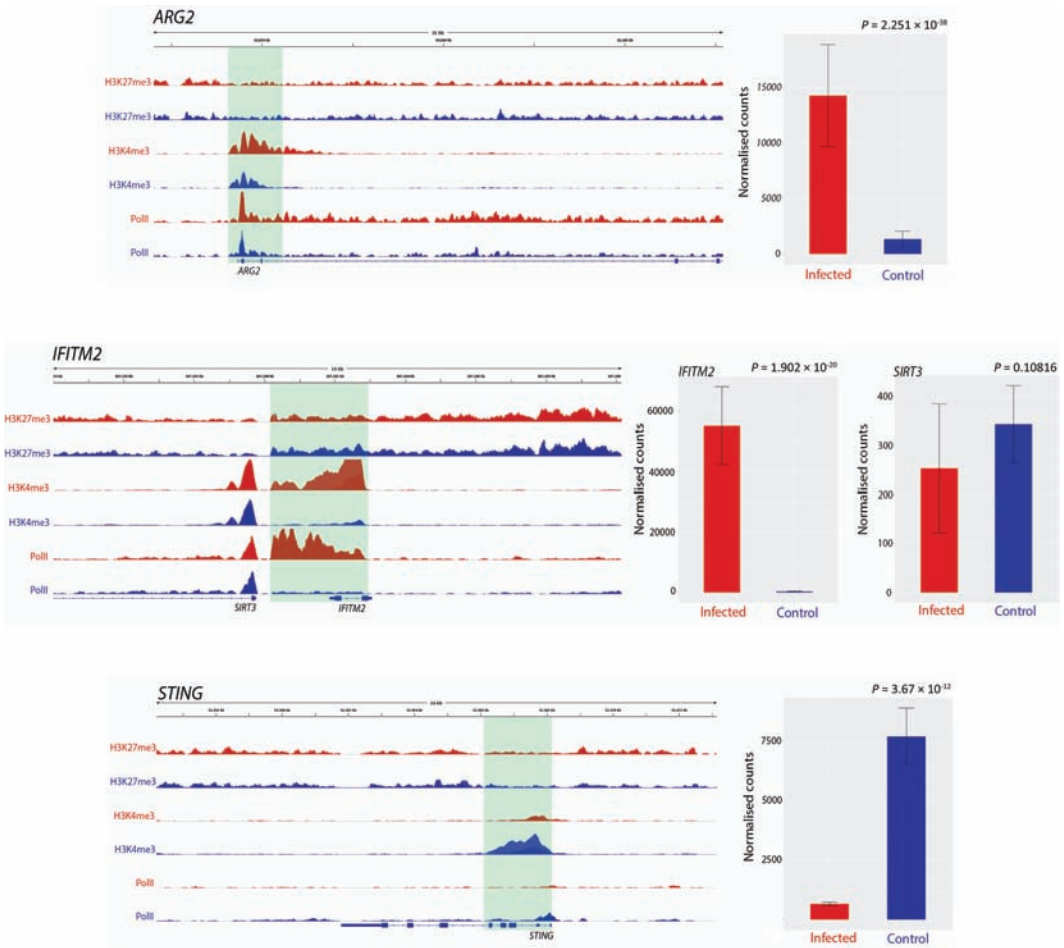


FIGURE 1 | Track visualization of *M. bovis* induced H3K4me3 and Pol II occupancy with relative change in expression at three immune response associated genes. Examples of signal tracks illustrating peaks of H3K27me3 (top two tracks), H3K4me3 (middle two tracks) and Pol II (bottom two tracks) in infected (red) and non-infected (blue) bovine alveolar macrophages, with the bovine reference genome on the bottom of each panel reading left to right. Accompanying each track image is the expression of the corresponding gene, with normalised counts of infected cells in red and control in blue. The *ARG2* gene exhibited an increase in H3K4me3 at 24 hpi as evidenced by the larger red H3K4me3 and red Pol II peaks. The *IFITM2* gene also exhibited larger H3K4me3 and Pol II peaks in infected samples; however, in contrast to this, *SIRT3*, which is located ~20 kb upstream from *IFITM2* gene, had no significant change in either peak. *TMEM173* (aka *STING*) exhibits an opposite pattern to most genes identified as having differential H3K4me3, where a larger peak is observed in control samples rather than infected.

affinity scores with MACS2 and diffBind, respectively; and 5) visual inspection of the tracks of the peaks confirmed the computationally determined differences in each condition.

Peaks that occurred in a sample indicate that H3K4me3 and Pol II are highly correlated with condition (**Figure 2A**); this demonstrates that the differences in H3 modifications are a result of infection rather than genomic differences between animals. **Figure 2B** further illustrates this, with the overlap in enriched peaks for H3K4me3 and Pol II being greater between condition than animal, i.e. the common number of H3K4me3 peaks between animal 1 control and animal 1 infected is 316 and the common number of H3K4me3 peaks between animal 1 infected and animal 2 infected is 798. **Figures 2C, D** illustrate that the distribution of the peaks, or sites with increased binding affinity, is differentially distributed between control and infected for both H3K4me3 and Pol II. Binding site affinity for H3K27me3 showed no significant differences between the control and infected groups for any genes. Analysis of genome-wide H3K4me3 revealed significant peak differences between control and infected samples at multiple sites in the genome under these criteria, with some of these differences occurring at the transcriptional start site of 233 genes. (**Figures 2A–D** and **Supplementary Figure 4**). **Supplementary Figure 1** demonstrates that the differences in H3K4me3 and Pol II peaks are minor, with cells from both conditions sharing most peaks and differing by only 1.8–2.95% in peaks across the genome. Principal component analysis (PCA) of the H3K4me3 mark and Pol II data indicated that these H3K4me3 and Pol II peak differences are strongly associated with *M. bovis* infection of bAM (**Supplementary Figure 3**).

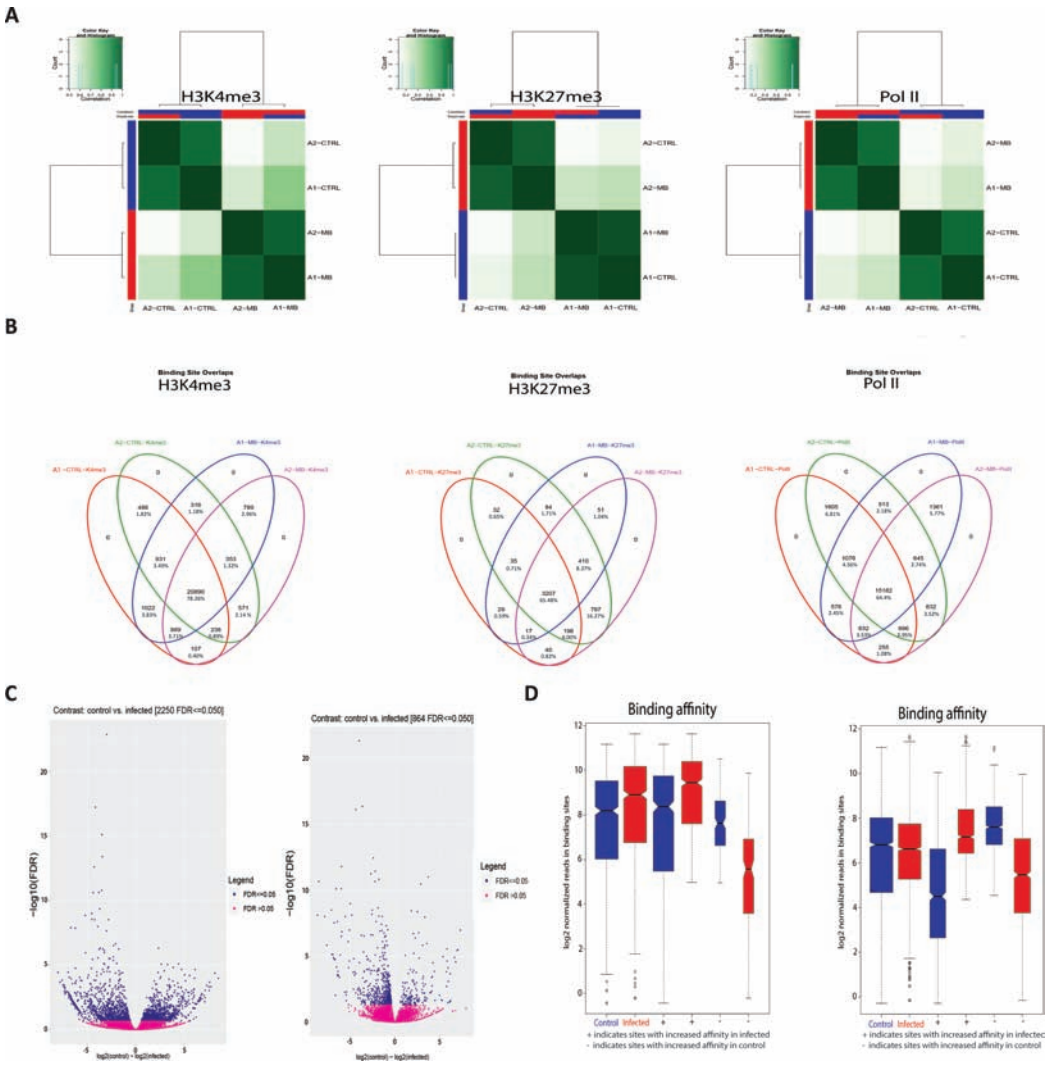


FIGURE 2 | *M. bovis* induced histone modifications occur genome wide at key immune loci. **(A)** Correlation heatmaps of differential peaks for H3K4me3, H3K27me3 and Pol II. Every peak location that is not consistent between each animal in each condition (i.e. a peak only occurs in the control group) is compared to determine if these inconsistent peaks are correlated with the animal or the condition. The differential peaks in H3K4me3 and Pol II correlate highly with condition, whereas there were no significant global differences in the distribution of H3K27me3. **(B)** Venn diagrams of differential peaks for H3K4me3, H3K27me3 and Pol II. Each condition shares most peaks. Where differences occur at TSS of genes, these genes are frequently associated with immune function. **(C)** Volcano plots of differential peaks for H3K4me3 and Pol II. The y-axis shows significance as FDR and the x-axis indicates increase in affinity for control (left) and infected (right). Significant sites are denoted in blue. **(D)** Boxplots of differential peaks for H3K4me3 and Pol II. Infected bAM are shown in red and control bAM are shown in blue. The left two boxes of each plot show distribution of reads over all differentially bound sites in the infected and control groups. The middle two boxes of each plot show the distribution of reads in differentially bound sites that increase in affinity in the control group. The far-right boxes in each plot show the distribution of reads in differentially bound sites that increase in affinity in the infected group.

Changes in H3K4me3 Are Accompanied by Immune Related Transcriptional Reprogramming

Previous studies have shown that increased H3K4me3 is frequently accompanied by an increase in Pol II occupancy and elevated expression of proximal genes (Clouaire et al., 2012; Barski et al., 2017). In the present study, we observed that H3K4me3 is accompanied by an increase in Pol II occupancy (**Figure 1** and **Supplementary Figure 4**). For a small number of genes (24 out of 233) where the H3K4me3 peak was larger in the control than the infected samples, Pol II occupancy was greater in control bAM for 20 genes (83.3%) and greater in infected bAM for 3 genes (12.5%). Conversely, where the H3K4me3 peak was larger in the infected bAM, Pol II occupancy was greater in the infected samples for 127 genes (60.4%) and greater in the control bAM for 14 genes (6.6%). The

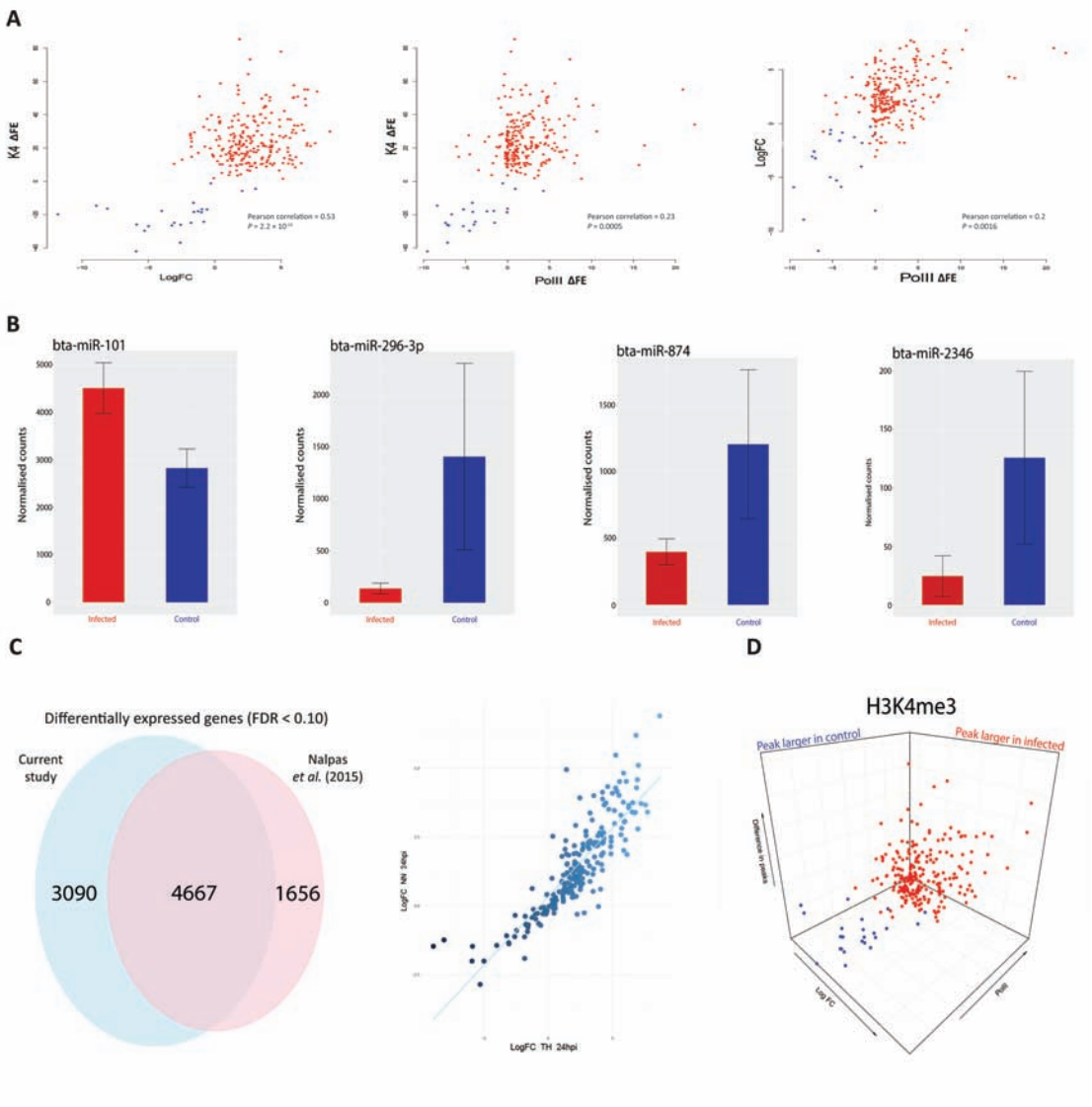


FIGURE 3 | H3K4me3 is accompanied by functional changes in Pol II occupancy, gene expression and gene regulation. **(A)** Scatter plots of H3K4me3 against Pol II occupancy and gene expression. The first plot is the difference of peaks for H3K4me3 between conditions. The values on the y-axis correspond to the difference in fold enrichment (ΔFE) of each peak at each gene between the control and infected groups. A positive value (x) means that the peak was x-fold enriched in the infected cells (red dots) compared to the control cells. Negative values correspond to the peaks being larger in the control (blue dots), indicating a decrease or total depletion of the peak in the infected cells. The x-axis represents the $\log_2$FC for each of the 232 genes, with each gene as a single data point. The second plot also has H3K4me3 on the y-axis but with peak differences in Pol II on the x-axis, with negative and positive values corresponding to greater occupancy in the control and infected samples, respectively. The final plot shows $\log_2$FC relative to Pol II occupancy. **(B)** Plots of normalised miRNA-seq counts. Each plot represents the normalised counts of a miRNA that was detected as exhibiting differential expression. Bta-miR-101 interacts with *ARG2*, bta-miR-296-3p with *TMEM173* (aka *STING*), bta-miR-874 with *BCL2A1* and bta-miR-2346 with *STAT1*. Red bars indicate infected and blue represent control samples. **(C)** Correlation and Venn diagram for both RNA-seq studies. The x-axis of the scatter plot represents the $\log_2$FC for each of the 232 genes from this study and the y-axis represents the $\log_2$FC for each of the 232 genes from the previous study (Nalpas et al., 2015). The Venn diagram shows the global overlap of differentially expressed genes from both studies with an FDR cut-off < 0.1. **(D)** 3-D plots for all three data sets. A combination of all three scatter plots from **Figure 3A**. Data points are genes. Blue genes are those that exhibited greater H3K4me3 in control bAM; red exhibited greater H3K4me3 in infected bAM.

remaining 60 genes (25%) did not exhibit H3K4me-associated Pol II occupancy in either control or infected samples. **Figure 3A** illustrates this trend, showing that Pol II occupancy normally accompanies H3K4me3.

To establish if H3K4me3 mark patterns were correlated with changes in gene expression, control non-infected bAM and bAM infected *M. bovis* AF2122/97 from four animals 24 hpi (including the two animals used for ChIP-seq) were used to generate eight

RNA-seq libraries. RNA-seq analysis revealed 7,757 differentially expressed genes ($log_2FC > 0$: 3,723 genes; $log_2FC < 0$: 4,034 genes; FDR < 0.1). Of the 233 genes identified in the ChIP-seq analysis, 232 (99.6%) were differentially expressed under these criteria (see **Supplementary Information File 2**). Of the genes that exhibited H3K4me3 peaks that were larger in the infected bAM, 21 (10%) were downregulated and 189 (90%) were upregulated. Of the genes that exhibited larger H3K4me3 peaks in the control group, 22 (91.6%) were downregulated and 2 (8.4%) were upregulated (**Figure 3A**). This pattern of directional gene expression correlating with H3K4me3 for the control and infected samples is consistent with the literature (Clouaire et al., 2012; Barski et al., 2017).

Existing published RNA-seq data generated by our group using *M. bovis*-infected ($n = 10$) and control non-infected bAM ($n = 10$) at 24 hpi (Nalpas et al., 2015) were also examined in light of the results from the present study. For the 232 genes identified here, a Pearson correlation coefficient of 0.85 was observed for two data sets (**Figure 3C**), thus demonstrating that gene expression differences between *M. bovis*-infected and control non-infected bAM are consistent across experiments, even where samples sizes are markedly different.

Transcriptional Reprogramming Is Coupled With Differential microRNA Expression

We have previously demonstrated that differential expression of immunoregulatory microRNAs (miRNAs) is evident in bAM infected with *M. bovis* compared to non-infected control bAM (Vegh et al., 2013; Vegh et al., 2015). To investigate the expression of miRNA in bAM used for the ChIP-seq analyses, miRNA was extracted and sequenced from the samples used for the RNA-seq analysis. Twenty-three differentially expressed miRNAs were detected at 24 hpi ($log_2FC > 0$: 13; $log_2FC < 0$: 10; FDR < 0.10). Of the 232 genes identified in the ChIP-seq/RNA-seq analysis, 93 are potential targets for the 23 differentially expressed miRNAs (**Supplementary Information File 3**). Further examination revealed that multiple immune genes, such as *BCL2A1* (bta-mir-874), *ARG2* (bta-mir-101), *TMEM173* (aka *STING*) (bta-mir-296-3p) and *STAT1* (bta-mir-2346), are potential regulatory targets for these miRNAs (**Figure 3B**). This observation therefore supports the hypothesis that miRNAs function in parallel with chromatin modifications to modulate gene expression in response to infection by *M. bovis*.

Integration of ChIP-Seq and RNA-Seq Data

The H3K4me3, Pol II, H3K27me3 ChIP-seq data and the RNA-seq data were subsequently integrated to evaluate the relationship between histone modifications and gene expression changes. Three-dimensional plots were generated to visualise the global differences between H3K4me3, Pol II and gene expression in infected and non-infected bAM (**Figure 3D**). These plots show that reduction of H3K4me3 in infected cells is associated with a decrease in gene expression and an absence of Pol II occupancy. Genome-wide H3K27me3 was also investigated to determine whether methylation of this residue was altered in response to *M. bovis* infection and if it was related to gene expression. No significant differences for H3K27me3 between control and infected bAM were detected, indicating that repression of gene expression through H3K27me3 does not play a role in the bAM response to *M. bovis* at 24 hpi. However, **Supplementary Figure 2** indicates that the presence of a H3K27me3 peak in both control and infected cells at the TSS of a H3K4me3 enriched gene correlated well with a lower or complete lack of Pol II occupancy.

Pathway Analysis Reveals H3K4me3 Marks Are Enriched for Key Immunological Genes

To identify biological pathways associated with genes identified through the ChIP-seq analyses, we integrated the ChIP-seq, RNA-seq and miRNA-seq data, which generated a panel of 93 genes that overlapped across each of the three data sets. Pathway analyses were carried out using three software tools: Ingenuity Pathway Analysis (IPA), Panther and DAVID (Thomas et al., 2003; Huang da et al., 2009; Kramer et al., 2014). IPA revealed an association with *respiratory illness* and the *innate immune response* (**Supplementary File 2**). Panther was used to examine the GO categories of the 232 genes (**Figure 4A**); this revealed enrichment for metabolic processes, response to stimuli and cellular processes, indicating that increased H3K4me3 in response to *M. bovis* infection occurs at TSS of genes associated with the immune response and at genes encoding key components of internal macrophage cellular regulation. GO unifies genes based on their gene and gene product attributes, which represents a useful method of identifying the families of gene functions for a given enriched gene set such as the one summarised in **Figure 4A** (The Gene Ontology Consortium, 2019).

The final part of the pathway analysis was performed using DAVID (Huang da et al., 2009). DAVID uses a list of background genes and query genes (in this case the 232 common genes across data sets) and identifies enriched groups of genes with shared biological functions. The DAVID analysis demonstrated that the 232 genes are involved in several signalling pathways, including the PI3K/AKT/mTOR, JAK-STAT and RIG-I-like signalling pathways (**Figure 4B** and the top 10 pathways are detailed in **Supplementary Information File 3**).

GWAS Integration Prioritises Bovine SNPs Associated With Resilience to *M. bovis* Infection

Previous work used high-density SNP (597,144 SNPs) data from 841 Holstein-Friesian bulls for a GWAS to detect SNPs associated with susceptibility/resistance to *M. bovis* infection (Richardson et al., 2016). Using a permutation-based approach to generate null SNP distributions, we leveraged these data to show that genomic regions within 100 kb up- and downstream of each of the 232 genes exhibiting differential H3K4me3 ChIP-seq peaks are significantly enriched for additional SNPs associated with resilience to *M. bovis* infection.

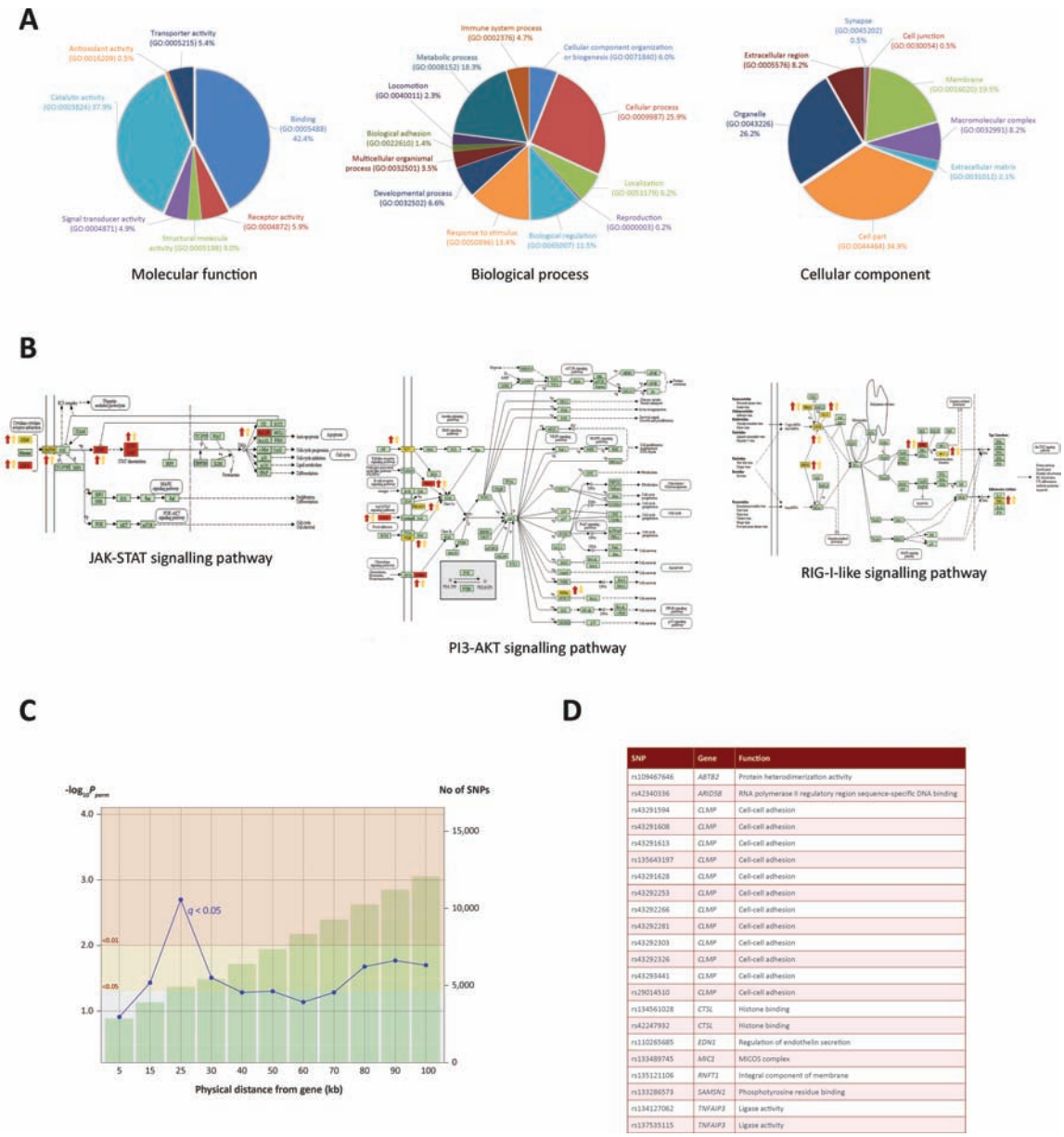


FIGURE 4 | Gene ontology enrichment and pathway analysis. **(A)** Gene ontology pie charts generated through PANTHER pathway analysis; 232 genes cluster by gene ontology under three main categories: *Biological process*, *Cellular component* and *Molecular function*. **(B)** KEGG pathway images containing genes identified from the ChIP-seq and RNA-seq analysis. Gene symbols coloured in yellow were identified in the ChIP-seq and RNA-seq analysis. Gene symbols coloured in red were also targeted by one or more differentially expressed miRNAs. Up or down red arrows indicate greater H3K4me3 in infected or control, respectively. Up or down yellow arrows indicate log2FC increase or decrease of the associated gene, respectively. **(C)** Line graph showing different genomic ranges from genes that are enriched for significant SNPs from GWAS data for bTB resilience. The bars represent the number of SNPs that occupy each range from each ChIP-seq enriched gene, with more SNPs correlating with a greater distance. The blue plotted line represents the negative $\log_{10}$ probability that the significant SNPs found at each distance at 0.05 FDR *q* value are significant by chance, with SNPs at 25 kb exhibiting the lowest probability. The null SNP *P* value distribution for each data point was generated from 1,000 permutations of random SNPs corresponding to the number of SNPs observed in a particular genomic range. **(D)** Genes enriched for SNPs significantly associated with resilience to *M. bovis* infection. SNP IDs and functional information obtained from the GeneCards® database (Stelzer et al., 2016) are also shown.

In total, 12,056 SNPs within the GWAS data set were located within 100 kb of the 232 H3K4me3 genes. Of these SNPs, up to 26 were found to be significantly associated with bTB susceptibility, depending on the distance interval of each gene. Interestingly, 22 SNPs found within 25 kb of 11 genes were found to be most significant at *P* and *q* values < 0.05, with declining significance of association as the region extended beyond 25 kb (**Figures 4C**, **4D** and **Supplementary File 3**). Significant SNPs were detected in

proximity to the following genes: *SAMSN1, CTSL, TNFAIP3, CLMP, ABTB2, RNFT1, MIC1, MIC2, EDN1* and *ARID5B*, all of which had significant differential enrichment of H3K4me3.

DISCUSSION

H3K4me3 Mark Occurs at Key Immune Genes

Our study has generated new information regarding host–pathogen interaction during the initial stages of *M. bovis* infection. We demonstrate that chromatin is remodelled through differential H3K4me3 and that Pol II occupancy is altered at key immune genes in *M. bovis*-infected bAM. This chromatin remodelling correlates with changes in the expression of genes that are pivotal for the innate immune response to mycobacteria (Nalpas et al., 2015; Alcaraz-Lopez et al., 2017; Malone et al., 2018). Our work supports the hypothesis that chromatin modifications of the host macrophage genome play an essential role during intracellular infections by mycobacterial pathogens (Cheng et al., 2014; LaMere et al., 2016).

The top pathways identified were the JAK-STAT signalling pathway, the PI3K/AKT/mTOR signalling pathway and the RIG-I-like receptor signalling pathway. In mammals, the JAK-STAT pathway is the principal signalling pathway that modulates expression of a wide array of cytokines and growth factors, involved in cell proliferation and apoptosis (Rawlings et al., 2004). The JAK-STAT signalling pathway and its regulators are also associated with coordinating an effective host response to mycobacterial infection (Manca et al., 2005; Cliff et al., 2015). Two JAK-STAT associated stimulating factors and a ligand receptor that exhibited increased H3K4me3 marks in infected samples were encoded by the *OSM*, *CSF3* and *CNTFR* genes, respectively (Marino and Roguin, 2008; Pastuschek et al., 2015). *OSM* has previously been shown to be upregulated in cells infected with either *M. bovis* or *M. tuberculosis* (O'Kane et al., 2008; Nalpas et al., 2015; Polena et al., 2016). Our work shows that this increased expression in response to mycobacteria is facilitated by H3K4me3-mediated chromatin accessibility. The protein encoded by *CSF3* has also been implicated as an immunostimulator in the response to mycobacterial infection due to its role in granulocyte and myeloid haematopoiesis (Martins et al., 2010). *CNTFR* encodes a ligand receptor that stimulates the JAK-STAT pathway and shows increased expression in other studies of mycobacterial infection (Nalpas et al., 2015; Malone et al., 2018). Following stimulation of JAK through ligand receptor binding, *STAT1* expression is increased. STAT1, a signal transducer and transcription activator that mediates cellular responses to interferons, cytokines and growth factors, is a pivotal JAK-STAT component and a core component in the response to mycobacterial infection (Tsumura et al., 2012). Here, the TSS of *STAT1* was associated with an increased deposition of H3K4me3. Interestingly, upregulation of *STAT1* was associated with a downregulation of bta-miR-2346, predicted to be a negative regulator of STAT1 (see **Supplementary Information File 3**). Overall, these results show that major components of the JAK-STAT pathway undergo chromatin remodelling mediated *via* H3K4me3, thereby facilitating activation and propagation of the JAK-STAT pathway through chromatin accessibility.

Key genes encoding components of the PI3K/AKT/mTOR pathway, such as *IRF7*, *RAC1* and *PIK3AP1*, were also identified as having increased H3K4me3 in *M. bovis* infected macrophages. PI3K/AKT/mTOR signalling contributes to a variety of processes that are critical in mediating aspects of cell growth and survival (Yu and Cui, 2016). Phosphatidylinositol-3 kinases (PI3Ks) and the mammalian target of rapamycin (mTOR) are integral to coordinating innate immune defences (Weichhart and Saemann, 2008). The PI3K/AKT/mTOR pathway is an important regulator of type I interferon production *via* activation of the interferon-regulatory factor 7, IRF7. RAC1 is a key activator of the PI3K/AKT/mTOR pathway and, in its active state, binds to a range of effector proteins to regulate cellular responses such as secretory processes, phagocytosis of apoptotic cells and epithelial cell polarisation (Yip et al., 2007). In addition, *in silico* analysis of our differentially expressed miRNAs predicted that several miRNAs, such as bta-miR-1343-3p, bta-miR-2411-3p and bta-miR-1296, regulate *RAC1*. *PIK3AP1* expression was also increased, in line with previous mycobacterial infection studies (Nalpas et al., 2015; Malone et al., 2018). Hence, as observed with the JAK-STAT pathway, H3K4me3 at these key PI3K/AKT/mTOR pathway genes acts to regulate the innate response to mycobacterial infection. In addition, the RIG-I-like receptor signalling pathway was also highlighted by the ChIP-seq, RNA-seq and miRNA-seq integrative analyses. Genes encoding multiple components of this pathway, such as *TRIM25*, *ISG15*, *IRF7* and *IKBKE*, were enriched for H3K4me3 and Pol II occupancy in *M. bovis*-infected bAM. The RIG-I-like receptor signalling pathway activates transcription factors that regulate production of type I interferons (Loo and Gale, 2011) and our results demonstrate that activation of this pathway in *M. bovis*-infected bAM is driven, to a large extent, by reconfiguration of the host chromatin.

H3K4me3 enriched loci are also flanked by genomic polymorphisms associated with resilience to *M. bovis* infection. Integration of our data with GWAS data from 841 bulls that have robust phenotypes for bTB susceptibility/resistance revealed 22 statistically significant SNPs within 25 kb of 11 H3K4me3 enriched genes. Statistical significance was determined if the newly permuted q values of every SNP found in proximity to each of the H3K4me3 enriched genes is unique to the observed set, when compared to 1,000 random sets of SNPs from the same GWAS (i.e. if significant q values of the same value or less occur with the same or greater frequency in randomised SNP sets, the observed SNPs are not deemed to be statistically significant). Most of these genes are involved in host immunity, with *CTSL, TNFAIP3* and *RNFT1* directly implicated in the human response to *M. tuberculosis* infection (Nepal et al., 2006; Silver et al., 2009; Meenu et al., 2016). The reprioritisation of genomic regions and array-based SNPs using integrative genomics approaches will be relevant for genomic prediction and genome-enabled breeding

and may facilitate fine-mapping efforts and the identification of targets for genome editing of cattle resilient to bTB.

H3K4me3 Deposition at Host Macrophage Genes and Immunological Evasion by *M. bovis*

The present study has revealed elevated H3K4me3 deposition and Pol II occupancy at key immune genes that are involved in the innate response to mycobacterial infection. In addition, we also identified several immune genes that had differential H3K4me3 and expression, where the expression change may be detrimental to the ability the host macrophage to clear infection. An example of this is *ARG2*, which exhibited increased H3K4me3 deposition, Pol II occupancy and expression ($Log_2FC = 3.415$, $P_{adj} = 7.52 \times 10^{-16}$) in infected cells. However, it is also interesting to note that the integrated expression output of *ARG2* may also be determined by the bta-miR-101 miRNA, a potential silencer of *ARG2* expression, which was observed to be upregulated in infected cells. Elevated levels of arginase 2, the protein product of the *ARG2* gene, have previously been shown to shift macrophages to an M2 phenotype (Lewis et al., 2011; Hardbower et al., 2016), which is anti-inflammatory and exhibits decreased responsiveness to IFN-γ and decreased bactericidal activity (Huang et al., 2015). Hence, it may be hypothesised that *M. bovis* infection triggers H3K4me3 deposition at the TSS of *ARG2* to drive an M2 phenotype and generate a more favourable niche for the establishment of infection. Like *ARG2,* increased expression of *BCL2A1* in *M. bovis*-infected bAM may also facilitate development of a replicative milieu for intracellular mycobacteria. Increased expression of *BCL2A1* is associated with decreased macrophage apoptosis (Vogler, 2012), which would otherwise restrict replication of intracellular pathogens.

In comparison to control non-infected bAM, the *TMEM173* (aka *STING*) gene exhibited substantially decreased expression in *M. bovis*-infected bAM ($Log_2FC = -3.225$, $P_{adj} = 8.64 \times 10^{-11}$). *TMEM173* encodes transmembrane protein 173, which drives interferon production and as such is a major regulator of the innate immune response to viral and bacterial infections, including *M. bovis* and *M. tuberculosis* (Manzanillo et al., 2012; McNab et al., 2015; Malone et al., 2018). Downregulation of *TMEM173* indicates that *M. bovis* can actively reduce or block methylation of H3K4 at this gene in infected macrophages, thereby enhancing intracellular survival of the pathogen. In this regard, we have recently shown that infection of bAM with *M. tuberculosis*, which is attenuated in cattle, causes increased *TMEM173* expression compared to infection with *M. bovis* (Malone et al., 2018).

The molecular mechanisms that pathogens employ to manipulate the host genome to subvert or evade the immune response are yet to be fully elucidated. Hijacking the host's own mechanisms for chromatin modulation is one potential explanation that has garnered attention in recent years (Hamon and Cossart, 2008; Rolando et al., 2015). These modulations of the host chromatin in bAM may be mediated through *M. bovis*-derived signals transmitted through bacterial metabolites, RNA-signalling or secreted peptides (Silmon de Monerri and Kim, 2014; Sharma et al., 2015; Yaseen et al., 2015; Woo and Alenghat, 2017).

CONCLUSIONS

Elucidation of the mechanisms used by pathogens to establish infection, and ultimately cause disease, requires an intimate knowledge of host–pathogen interactions. Using transcriptomics and epigenomics, we have identified altered expression of major host immune genes following infection of primary bovine macrophages with *M. bovis*. We have shown that reprogramming of the alveolar macrophage transcriptome occurs mainly through increased deposition of H3K4me3 at key immune function genes, with additional gene expression modulation *via* miRNA differential expression. This modulation of gene expression drives a shift of the macrophage phenotype towards the more replication-permissive M2 macrophage phenotype. We have also identified that alveolar macrophages infected with *M. bovis* exhibit differentially expressed genes (in regions with modified chromatin) that are enriched for significant SNPs from GWAS data for bTB resilience. Finally, our results support the emerging concept that pathogens can hijack host chromatin, through manipulation of H3K4me3, to subvert host immunity and to establish infection.

DATA AVAILABILITY STATEMENT

The ChIP-seq, RNA-seq and microRNA-seq data sets have been submitted to the NCBI Gene Expression Omnibus (GEO) with accession number GSE116734.

ETHICS STATEMENT

All animal procedures were performed according to the provisions of Statutory Instrument No. 543/2012 (under Directive 2010/63/EU on the Protection of Animals used for Scientific Purposes). Ethical approval was obtained from the University College Dublin Animal Ethics Committee (protocol number AREC-13-14-Gordon).

AUTHOR CONTRIBUTIONS

Project conceptualisation: AO'D, DM and TH. Software and formal analysis: TH and MM. Investigation: AO'D, DV and JB. Resources: DM, SG and DV. Data curation: TH. Writing—original draft: TH, AO'D and DM. Writing—review and editing: D.V., SG and MM. Supervision and project administration: DM and AO'D. Funding acquisition: DM, SG, DV and AO'D.

FUNDING

This study was supported by Science Foundation Ireland (SFI) Investigator Programme Awards to DM and SG (grant nos. SFI/08/IN.1/B2038 and SFI/15/IA/3154); a European Union Framework 7 Project Grant to DM (no: KBBE-211602-MACROSYS); an EU H2020 COST Action short-term scientific mission (STSM) grant to AO'D (reference code: COST-STSM-ECOST-STSM-CA15112-050317-081648); a University of Edinburgh Chancellor's Fellowship to DV; and BBSRC Institute Strategic Grant funding to the Roslin Institute (grant nos. BBS/E/D/10002070 and BBS/E/D/20002172). The funding agencies had no role in the study design, collection, analysis and interpretation of data, and no role in writing the manuscript.

ACKNOWLEDGMENTS

The authors would also like to thank FAANG–Europe for awarding AO'D a short-term scientific mission (STSM) grant. We would also like to acknowledge Edinburgh Genomics for generation of sequencing data. This manuscript has been released as a pre-print at *bioRχiv* (Hall et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01386/full#supplementary-material

REFERENCES

Alcaraz-Lopez, O. A., Garcia-Gil, C., Morales-Martinez, C., Lopez-Rincon, G., Estrada-Chavez, C., Gutierrez-Pabello, J. A., et al. (2017). Divergent macrophage responses to *Mycobacterium bovis* among naturally exposed uninfected and infected cattle. *Immunol. Cell Biol.* 95, 436–442. doi: 10.1038/icb.2016.114

Andrews, S. (2016). FastQC: a quality control tool for high throughput sequence data. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arts, R. J. W., Moorlag, S., Novakovic, B., Li, Y., Wang, S. Y., Oosting, M., et al. (2018). BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host Microbe.* 23, 89–100 e105. doi: 10.1016/j.chom.2017.12.010

Awuh, J. A., and Flo, T. H. (2017). Molecular basis of mycobacterial survival in macrophages. *Cell Mol Life Sci* 74, 1625–1648. doi: 10.1007/s00018-016-2422-8

Barski, A., Cuddapah, S., Kartashov, A. V., Liu, C., Imamichi, H., Yang, W., et al. (2017). Rapid recall ability of memory T cells is encoded in their epigenome. *Sci. Rep.* 7, 39785. doi: 10.1038/srep39785

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Method* 57, 289–300.

Bierne, H., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2, a010272. doi: 10.1101/cshperspect.a010272

BoseDasgupta, S., and Pieters, J. (2018). Macrophage-microbe interaction: lessons learned from the pathogen *Mycobacterium tuberculosis*. *Semin. Immunopathol.* 40, 577–591. doi: 10.1007/s00281-018-0710-0

Bouttier, M., Laperriere, D., Memari, B., Mangiapane, J., Fiore, A., Mitchell, E., et al. (2016). Alu repeats as transcriptional regulatory platforms in macrophage responses to *M. tuberculosis* infection. *Nucleic. Acids Res.* 44, 10571–10587. doi: 10.1093/nar/gkw782

Buddle, B. M., Vordermeier, H. M., and Hewinson, R. G. (2016). Experimental infection models of tuberculosis in domestic livestock. *Microbiol Spectr.* 4. doi: 10.1128/microbiolspec.TBTB2-0017-2016

Cambier, C. J., Falkow, S., and Ramakrishnan, L. (2014). Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. *Cell* 159, 1497–1509. doi: 10.1016/j.cell.2014.11.024

Chaurasiya, S. K. (2018). Tuberculosis: Smart manipulation of a lethal host. *Microbiol. Immunol.* 62, 361–379. doi: 10.1111/1348-0421.12593

Cheng, J., Blum, R., Bowman, C., Hu, D., Shilatifard, A., Shen, S., et al. (2014). A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol. Cell* 53, 979–992. doi: 10.1016/j.molcel.2014.02.032

Cliff, J. M., Kaufmann, S. H., McShane, H., van Helden, P., and O'Garra, A. (2015). The human immune response to tuberculosis and its treatment: a view from the blood. *Immunol. Rev.* 264, 88–102. doi: 10.1111/imr.12269

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., et al. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* 26, 1714–1728. doi: 10.1101/gad.194209.112

de Chastellier, C. (2009). The many niches and strategies used by pathogenic mycobacteria for survival within host macrophages. *Immunobiology* 214, 526–542. doi: 10.1016/j.imbio.2008.12.005

Ehrt, S., and Schnappinger, D. (2009). Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol.* 11, 1170–1178. doi: 10.1111/j.1462-5822.2009.01335.x

Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from ChIP-Seq data. *Curr. Protoc. Bioinformatics* 2, 14. doi: 10.1002/0471250953.bi0214s34

Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., et al. (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7877–7882. doi: 10.1073/pnas.1130426100

Hall, T. J., Vernimmen, D., Browne, J. A., Mullen, M. P., Gordon, S. V., MacHugh, D. E., et al. (2019). Alveolar macrophage chromatin is modified to orchestrate host response to *Mycobacterium bovis* infection. *bioRxiv*, 520098. doi: 10.1101/520098

Hamon, M. A., and Cossart, P. (2008). Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe.* 4, 100–109. doi: 10.1016/j.chom.2008.07.009

Hardbower, D. M., Asim, M., Murray-Stewart, T., Casero, R. A.Jr., Verriere, T., Lewis, N. D., et al. (2016). Arginase 2 deletion leads to enhanced M1 macrophage activation and upregulated polyamine metabolism in response to *Helicobacter pylori* infection. *Amino Acids* 48, 2375–2388. doi: 10.1007/s00726-016-2231-2

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang, Z., Luo, Q., Guo, Y., Chen, J., Xiong, G., Peng, Y., et al. (2015). *Mycobacterium tuberculosis*-induced polarization of human macrophage orchestrates the formation and development of tuberculous granulomas *in vitro*. *PLoS ONE* 10, e0129744. doi: 10.1371/journal.pone.0129744

Hussain Bhat, K., and Mukhopadhyay, S. (2015). Macrophage takeover and the host-bacilli interplay during tuberculosis. *Future Microbiol.* 10, 853–872. doi: 10.2217/fmb.15.11

Jensen, K., Gallagher, I. J., Johnston, N., Welsh, M., Skuce, R., Williams, J. L., et al. (2018). Variation in the early host-pathogen interaction of bovine macrophages with divergent *Mycobacterium bovis* strains in the United Kingdom. *Infect Immun.* 86. doi: 10.1128/iai.00385-17

Kaufmann, S. H. E., and Dorhoi, A. (2016). Molecular determinants in phagocyte-bacteria interactions. *Immunity* 44, 476–491. doi: 10.1016/j.immuni.2016.02.014

Kramer, A., Green, J., Pollard, J.Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703

LaMere, S. A., Thompson, R. C., Komori, H. K., Mark, A., and Salomon, D. R. (2016). Promoter H3K4 methylation dynamically reinforces activation-

induced pathways in human CD4 T cells. *Genes Immun.* 17, 283–297. doi: 10.1038/gene.2016.19

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831. doi: 10.1101/gr.136184.111

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lavalett, L., Rodriguez, H., Ortega, H., Sadee, W., Schlesinger, L. S., and Barrera, L. F. (2017). Alveolar macrophages from tuberculosis patients display an altered inflammatory gene expression profile. *Tuberculosis* 107, 156–167. doi: 10.1016/j.tube.2017.08.012

Lerdrup, M., Johansen, J. V., and Agrawal-Singh, S. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat. Struct. Mol. Biol.* 23, 349–357. doi: 10.1038/nsmb.3180

Lewis, N. D., Asim, M., Barry, D. P., de Sablet, T., Singh, K., Piazuelo, M. B., et al. (2011). Immune evasion by *Helicobacter pylori* is mediated by induction of macrophage arginase II. *J. Immunol.* 186, 3632–3641. doi: 10.4049/jimmunol.1003431

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Loo, Y. M., and Gale, M.Jr. (2011). Immune signaling by RIG-I-like receptors. *Immunity* 34, 680–692. doi: 10.1016/j.immuni.2011.05.003

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Magee, D. A., Conlon, K. M., Nalpas, N. C., Browne, J. A., Pirson, C., Healy, C., et al. (2014). Innate cytokine profiling of bovine alveolar macrophages reveals commonalities and divergence in the response to *Mycobacterium bovis* and *Mycobacterium tuberculosis* infection. *Tuberculosis (Edinb)* 94, 441–450. doi: 10.1016/j.tube.2014.04.004

Malone, K. M., Rue-Albrecht, K., Magee, D. A., Conlon, K., Schubert, O. T., Nalpas, N. C., et al. (2018). Comparative 'omics analyses differentiate *Mycobacterium tuberculosis* and *Mycobacterium bovis* and reveal distinct macrophage responses to infection with the human and bovine tubercle bacilli. *Microb. Genom.* 4. doi: 10.1099/mgen.0.000163

Manca, C., Tsenova, L., Freeman, S., Barczak, A. K., Tovey, M., Murray, P. J., et al. (2005). Hypervirulent *M. tuberculosis* W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. *J. Interferon. Cytokine Res.* 25, 694–701. doi: 10.1089/jir.2005.25.694

Manzanillo, P. S., Shiloh, M. U., Portnoy, D. A., and Cox, J. S. (2012). *Mycobacterium tuberculosis* activates the DNA-dependent cytosolic surveillance pathway within macrophages. *Cell Host Microbe.* 11, 469–480. doi: 10.1016/j.chom.2012.03.007

Marino, V. J., and Roguin, L. P. (2008). The granulocyte colony stimulating factor (G-CSF) activates Jak/STAT and MAPK pathways in a trophoblastic cell line. *J. Cell Biochem.* 103, 1512–1523. doi: 10.1002/jcb.21542

Marr, A. K., MacIsaac, J. L., Jiang, R., Airo, A. M., Kobor, M. S., and McMaster, W. R. (2014). *Leishmania donovani* infection causes distinct epigenetic DNA methylation changes in host macrophages. *PLoS Pathog.* 10, e1004419. doi: 10.1371/journal.ppat.1004419

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 3. doi: 10.14806/ej.17.1.200

Martins, A., Han, J., and Kim, S. O. (2010). The multifaceted effects of granulocyte colony-stimulating factor in immunomodulation and potential roles in intestinal immune homeostasis. *IUBMB Life* 62, 611–617. doi: 10.1002/iub.361

McNab, F., Mayer-Barber, K., Sher, A., Wack, A., and O'Garra, A. (2015). Type I interferons in infectious disease. *Nat. Rev. Immunol.* 15, 87–103. doi: 10.1038/nri3787

Meenu, S., Thiagarajan, S., Ramalingam, S., Michael, A., and Ramalingam, S. (2016). Modulation of host ubiquitin system genes in human endometrial cell line infected with *Mycobacterium tuberculosis*. *Med. Microbiol. Immunol. (Berl)* 205, 163–171. doi: 10.1007/s00430-015-0432-z

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45, D183–D189. doi: 10.1093/nar/gkw1138

Nalpas, N. C., Magee, D. A., Conlon, K. M., Browne, J. A., Healy, C., McLoughlin, K. E., et al. (2015). RNA sequencing provides exquisite insight into the manipulation of the alveolar macrophage by tubercle bacilli. *Sci. Rep.* 5, 13629. doi: 10.1038/srep13629

Nepal, R. M., Mampe, S., Shaffer, B., Erickson, A. H., and Bryant, P. (2006). Cathepsin L maturation and activity is impaired in macrophages harboring *M. avium*, and *M. tuberculosis*. *Int. Immunol.* 18, 931–939. doi: 10.1093/intimm/dxl029

Niller, H. H., and Minarovits, J. (2016). Patho-epigenetics of infectious diseases caused by intracellular bacteria. *Adv. Exp. Med. Biol.* 879, 107–130. doi: 10.1007/978-3-319-24738-0_6

O'Doherty, A. M., O'Shea, L. C., and Fair, T. (2012). Bovine DNA methylation imprints are established in an oocyte size-specific manner, which are coordinated with the expression of the DNMT3 family proteins. *Biol. Reprod.* 86, 67. doi: 10.1095/biolreprod.111.094946

O'Kane, C. M., Elkington, P. T., and Friedland, J. S. (2008). Monocyte-dependent oncostatin M and TNF-alpha synergize to stimulate unopposed matrix metalloproteinase-1/3 secretion from human lung fibroblasts in tuberculosis. *Eur. J. Immunol.* 38, 1321–1330. doi: 10.1002/eji.200737855

Olea-Popelka, F., Muwonge, A., Perera, A., Dean, A. S., Mumford, E., Erlacher-Vindel, E., et al. (2017). Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis* – a call for action. *Lancet Infect. Dis.* 17, e21–e25. doi: 10.1016/S1473-3099(16)30139-6

Papp, A. C., Azad, A. K., Pietrzak, M., Williams, A., Handelman, S. K., Igo, R. P.Jr., et al. (2018). AmpliSeq transcriptome analysis of human alveolar and monocyte-derived macrophages over time in response to *Mycobacterium tuberculosis* infection. *PLoS ONE* 13, e0198221. doi: 10.1371/journal.pone.0198221

Pastuschek, J., Poetzsch, J., Morales-Prieto, D. M., Schleussner, E., Markert, U. R., and Georgiev, G. (2015). Stimulation of the JAK/STAT pathway by LIF and OSM in the human granulosa cell line COV434. *J. Reprod. Immunol.* 108, 48–55. doi: 10.1016/j.jri.2015.03.002

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Polena, H., Boudou, F., Tilleul, S., Dubois-Colas, N., Lecointe, C., Rakotosamimanana, N., et al. (2016). *Mycobacterium tuberculosis* exploits the formation of new blood vessels for its dissemination. *Sci. Rep.* 6, 33162. doi: 10.1038/srep33162

Queval, C. J., Brosch, R., and Simeone, R. (2017). The macrophage: A disputed fortress in the battle against *Mycobacterium tuberculosis*. *Front. Microbiol.* 8, 2284. doi: 10.3389/fmicb.2017.02284

Rawlings, J. S., Rosler, K. M., and Harrison, D. A. (2004). The JAK/STAT signaling pathway. *J. Cell Sci.* 117, 1281–1283. doi: 10.1242/jcs.00963

Richardson, I. W., Berry, D. P., Wiencko, H. L., Higgins, I. M., More, S. J., McClure, J., et al. (2016). A genome-wide association study for genetic susceptibility to *Mycobacterium bovis* infection in dairy cattle identifies a susceptibility QTL on chromosome 23. *Genet. Sel. Evol.* 48, 19. doi: 10.1186/s12711-016-0197-x

Rolando, M., Gomez-Valero, L., and Buchrieser, C. (2015). Bacterial remodelling of the host epigenome: functional role and evolution of effectors methylating host histones. *Cell. Microbiol.* 17, 1098–1107. doi: 10.1111/cmi.12463

Schorey, J. S., and Schlesinger, L. S. (2016). Innate immune responses to tuberculosis. *Microbiol. Spectr.* 4. doi: 10.1128/microbiolspec.TBTB2-0010-2016

Sharma, G., Upadhyay, S., Srilalitha, M., Nandicoori, V. K., and Khosla, S. (2015). The interaction of mycobacterial protein Rv2966c with host chromatin is mediated through non-CpG methylation and histone H3/H4 binding. *Nucleic Acids Res.* 43, 3922–3937. doi: 10.1093/nar/gkv261

Sharma, G., Sowpati, D. T., Singh, P., Khan, M. Z., Ganji, R., Upadhyay, S., et al. (2016). Genome-wide non-CpG methylation of the host genome during *M. tuberculosis* infection. *Sci. Rep.* 6, 25006. doi: 10.1038/srep25006

Silmon de Monerri, N. C., and Kim, K. (2014). Pathogens hijack the epigenome: a new twist on host-pathogen interactions. *Am. J. Pathol.* 184, 897–911. doi: 10.1016/j.ajpath.2013.12.022

Silver, R. F., Walrath, J., Lee, H., Jacobson, B. A., Horton, H., Bowman, M. R., et al. (2009). Human alveolar macrophage gene responses to *Mycobacterium*

tuberculosis strains H37Ra and H37Rv. *Am. J. Respir. Cell Mol. Biol.* 40, 491–504. doi: 10.1165/rcmb.2008-0219OC

Sims, R. J.3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. *Trends Genet.* 19, 629–639. doi: 10.1016/j.tig.2003.09.007

Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000.Res* 4, 1521. doi: 10.12688/f1000research.7563.2.

Stark, R., and Brown, G. (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

Steele, J. H. (1995). "Introduction (Part 2 Regional and Country Status Reports)," in *Mycobacterium bovis infection in animals and humans*. Eds. C. O. Thoen and J. H. Steele (Ames, IA, USA: Iowa State University Press), 169–172.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* 54, 1.30.31–31.30.33. doi: 10.1002/cpbi.5

Stutz, M. D., Clark, M. P., Doerflinger, M., and Pellegrini, M. (2018). *Mycobacterium tuberculosis*: Rewiring host cell signaling to promote infection. *J. Leukoc. Biol.* 103, 259–268. doi: 10.1002/JLB.4MR0717-277R

The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–d338. doi: 10.1093/nar/gky1055

Thoen, C. O., Kaplan, B., Thoen, T. C., Gilsdorf, M. J., and Shere, J. A. (2016). Zoonotic tuberculosis. A comprehensive ONE HEALTH approach. *Medicina (B Aires)* 76, 159–165.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141. doi: 10.1101/gr.772403

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017

Tsumura, M., Okada, S., Sakai, H., Yasunaga, S., Ohtsubo, M., Murata, T., et al. (2012). Dominant-negative STAT1 SH2 domain mutations in unrelated patients with Mendelian susceptibility to mycobacterial disease. *Hum. Mutat.* 33, 1377–1387. doi: 10.1002/humu.22113

Vayr, F., Martin-Blondel, G., Savall, F., Soulat, J. M., Deffontaines, G., and Herin, F. (2018). Occupational exposure to human *Mycobacterium bovis* infection: A systematic review. *PLoS Negl. Trop. Dis.* 12, e0006208. doi: 10.1371/journal.pntd.0006208

Vegh, P., Foroushani, A. B., Magee, D. A., McCabe, M. S., Browne, J. A., Nalpas, N. C., et al. (2013). Profiling microRNA expression in bovine alveolar macrophages using RNA-seq. *Vet. Immunol. Immunopathol.* 155, 238–244. doi: 10.1016/j.vetimm.2013.08.004

Vegh, P., Magee, D. A., Nalpas, N. C., Bryan, K., McCabe, M. S., Browne, J. A., et al. (2015). MicroRNA profiling of the bovine alveolar macrophage response to *Mycobacterium bovis* infection suggests pathogen survival is enhanced by microRNA regulation of endocytosis and lysosome trafficking. *Tuberculosis* 95, 60–67. doi: 10.1016/j.tube.2014.10.011

Vejnar, C. E., and Zdobnov, E. M. (2012). MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* 40, 11673–11683. doi: 10.1093/nar/gks901

Vernimmen, D., Lynch, M. D., De Gobbi, M., Garrick, D., Sharpe, J. A., Sloane-Stanley, J. A., et al. (2011). Polycomb eviction as a new distant enhancer function. *Genes. Dev.* 25, 1583–1588. doi: 10.1101/gad.16985411

Vogler, M. (2012). BCL2A1: the underdog in the BCL2 family. *Cell Death Differ.* 19, 67–74. doi: 10.1038/cdd.2011.158

Waters, W. R., Palmer, M. V., Buddle, B. M., and Vordermeier, H. M. (2012). Bovine tuberculosis vaccine research: historical perspectives and recent advances. *Vaccine* 30, 2611–2622. doi: 10.1016/j.vaccine.2012.02.018

Waters, W. R., Maggioli, M. F., McGill, J. L., Lyashchenko, K. P., and Palmer, M. V. (2014). Relevance of bovine tuberculosis research to the understanding of human disease: historical perspectives, approaches, and immunologic mechanisms. *Vet. Immunol. Immunopathol.* 159, 113–132. doi: 10.1016/j.vetimm.2014.02.009

Weichhart, T., and Saemann, M. D. (2008). The PI3K/Akt/mTOR pathway in innate immune cells: emerging therapeutic applications. *Ann. Rheum. Dis.* 67 Suppl 3, iii70–iii74. doi: 10.1136/ard.2008.098459

Weiss, G., and Schaible, U. E. (2015). Macrophage defense mechanisms against intracellular bacteria. *Immunol. Rev.* 264, 182–203. doi: 10.1111/imr.12266

Wilbanks, E. G., and Facciotti, M. T. (2010). Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS ONE* 5, e11471. doi: 10.1371/journal.pone.0011471

Williams, A., and Orme, I. M. (2016). Animal models of tuberculosis: an overview. *Microbiol. Spectr.* 4. doi: 10.1128/microbiolspec.TBTB2-0004-2015

Woo, V., and Alenghat, T. (2017). Host-microbiota interactions: epigenomic regulation. *Curr. Opin. Immunol.* 44, 52–60. doi: 10.1016/j.coi.2016.12.001

Yaseen, I., Kaur, P., Nandicoori, V. K., and Khosla, S. (2015). Mycobacteria modulate host epigenetic machinery by Rv1988 methylation of a non-tail arginine of histone H3. *Nat. Commun.* 6, 8922. doi: 10.1038/ncomms9922

Yip, S. C., El-Sibai, M., Coniglio, S. J., Mouneimne, G., Eddy, R. J., Drees, B. E., et al. (2007). The distinct roles of Ras and Rac in PI 3-kinase-dependent protrusion during EGF-stimulated cell migration. *J. Cell Sci.* 120, 3138–3146. doi: 10.1242/jcs.005298

Yu, J. S., and Cui, W. (2016). Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. *Development* 143, 3050–3060. doi: 10.1242/dev.137075

Zheng, L., Leung, E. T. Y., Wong, H. K., Lui, G., Lee, N., To, K.-F., et al. (2016). Unraveling methylation changes of host macrophages in *Mycobacterium tuberculosis* infection. *Tuberculosis* 98, 139–148. doi: 10.1016/j.tube.2016.03.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hall, Vernimmen, Browne, Mullen, Gordon, MacHugh and O'Doherty. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fgene.2019.01347

Emerging Roles of Heat-Induced circRNAs Related to Lactogenesis in Lactating Sows

Jiajie Sun[1†], Haojie Zhang[1†], Baoyu Hu[1], Yueqin Xie[1], Dongyang Wang[1], Jinzhi Zhang[2], Ting Chen[1], Junyi Luo[1], Songbo Wang[1], Qinyan Jiang[1], Qianyun Xi[1], Zujing Chen[1] and Yongliang Zhang[1*]*

[1] College of Animal Science, Guangdong Provincial Key Laboratory of Animal Nutrition Control, Guangdong Engineering & Research Center for Woody Fodder Plants, National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangzhou, China, [2] College of Animal Science, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
David E. MacHugh,
University College Dublin, Ireland

Reviewed by:
Gary Rohrer,
United States Department of Agriculture, United States
Nicolas Nalpas,
University of Tübingen, Germany

***Correspondence:**
Zujing Chen
zujingchen@scau.edu.cn
Yongliang Zhang
zhangyl@scau.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 29 May 2019
Accepted: 10 December 2019
Published: 11 February 2020

Citation:
Sun J, Zhang H, Hu B, Xie Y, Wang D, Zhang J, Chen T, Luo J, Wang S, Jiang Q, Xi Q, Chen Z and Zhang Y (2020) Emerging Roles of Heat-Induced circRNAs Related to Lactogenesis in Lactating Sows. Front. Genet. 10:1347. doi: 10.3389/fgene.2019.01347

Heat stress negatively influences milk production and disrupts normal physiological activity of lactating sows, but the precious mechanisms by which hyperthermia adversely affects milk synthesis in sows still remain for further study. Circular RNAs are a novel class of non-coding RNAs with regulatory functions in various physiological and pathological processes. The expression profiles and functions of circRNAs of sows in lactogenesis remain largely unknown. In the present study, long-term heat stress (HS) resulted in a greater concentration of serum HSP70, LDH, and IgG, as well as decreased levels of COR, SOD, and PRL. HS reduced the total solids, fat, and lactose of sow milk, and HS significantly depressed CSNαs1, CSNαs2, and CSNκ biosynthesis. Transcriptome sequencing of lactating porcine mammary glands identified 42 upregulated and 25 downregulated transcripts in HS vs. control. Functional annotation of these differentially-expressed transcripts revealed four heat-induced genes involved in lactation. Moreover, 29 upregulated and 21 downregulated circRNA candidates were found in response to HS. Forty-two positively correlated circRNA-mRNA expression patterns were constructed between the four lactogenic genes and differentially expressed circRNAs. Five circRNA-miRNA-mRNA post-transcriptional networks were identified involving genes in the HS response of lactating sows. In this study we establish a valuable resource for circRNA biology in sow lactation. Analysis of a circRNA-miRNA-mRNA network further uncovered a novel layer of post-transcriptional regulation that could be used to improve sow milk production.

Keywords: heat stress, lactating sow, circRNA, ceRNA, casein

INTRODUCTION

In seasonal climates, high ambient temperature is the primary environmental stress impacting domestic animal performance, including growth, reproduction, and lactation (Das et al., 2016). In general, high-yielding animals are especially susceptible to thermal stress since they generate considerably more metabolic heat (Kadzere et al., 2002). In response to heat stress (HS), dairy

animals experience a sustained reduction in appetite and nutrient uptake (Bohmanova et al., 2007) and a subsequent reallocation of energy for heat acclimation (Renaudeau et al., 2012), thereby resulting in decreased milk yield and milk quality, which negatively affects the efficiency and profitability of animal farms worldwide (Hill and Wall, 2015).

In modern swine husbandry, lactating sows have been heavily selected for increased productivity (fertility, disease resistance, feed conversion efficiency, and so on) during the last two decades, and are especially at risk of HS (Renaudeau, 2005), as they have a thermal neutral zone between 16 and 22°C (Messias de Bragança et al., 1998). It is noteworthy that high temperatures above 25° are sustained for half a year in the south of China; thus, local sows are often exposed to hot conditions. Under thermal stress, individuals normally increase respiration rates and reduce feed intake (Quiniou and Noblet, 1999), in an effort to generate a negative energy balance to promote metabolic heat loss to counter HS (Renaudeau et al., 2001; Renaudeau et al., 2012). In addition, HS also influences milk production in lactating sows, perhaps through an indirect effect associated with reduction in feed intake (Ribeiro et al., 2018); however, previous reports of Messias de Bragança et al. (1998) and Silanikove et al. (2009) suggested that there may be a direct effect of ambient temperature on mammary gland metabolism in connection with low milk yield. Thus, identifying key differences in the mammary gland of lactating sows in response to high ambient temperatures has the potential to improve the productivity of sows in adverse environments (Collier et al., 2006). The ability to use powerful genomic tools to evaluate genetic differences associated with thermal tolerance can provide important information on the underlying mechanisms of HS on lactation, and will permit the selection of sows for resistance to HS.

Circular RNAs (circRNAs) are a recently identified genetic element that are abundantly expressed, highly conserved between different animal species (Hanan et al., 2017), and are involved in the foundation of mammary gland growth and development (Zhang et al., 2015; Zhu et al., 2016), milk synthesis (Zhang C. et al., 2016), and secretion and transportation (Wang et al., 2019). HS greatly impacts circRNA biogenesis, and heat-induced circRNAs perform substantial regulatory functions through circRNA-mediated competing endogenous RNA (ceRNA) networks (Pan et al., 2018). Although patterns of circRNA expression and function have been revealed among various developmental stages and physiological conditions (Lai et al., 2018; Patop and Kadener, 2018), little is known about how HS affects circRNAs in lactation. In this study, we focused on circRNAs involved in the HS response of lactating sows, and we explored potential mechanisms underlying circRNA regulation in mammary tissue.

MATERIALS AND METHODS

Study Design and Sample Collection

A total of 60 healthy purebred Landrace sows (2–3 parity and without genetic relationships) were separated into two balanced cohorts of 30 animals each, and HS tests were conducted at a local thoroughbred farm during December 2016 and during August 2017 (WENS Shuitai Breeding Pig Farm, Guangdong, China). All sows were fed the same commercial formula diet and raised under the same management conditions. In the experimental stage, the ambient temperatures and relative humidity were measured at 14:00pm in everyday. In details, one cohort of 30 sows was selected in the winter months with a moderate average temperature, designated as the non-heat stress (NS) population; other cohort of 30 animals was selected in the summer months with a higher average temperature, designated as the HS population. Within each cohort, the number of piglets born alive was recorded, and litter birth weights of piglets were obtained within 24 hours after farrowing. Piglets were not offered creep feed, and sow milk was the only feed available to the piglets during lactation. On day 21, weaning survival and the weights of living piglets per litter were recorded and used to calculate average daily weight gain. Blood samples (10 ml) were collected at 10:00am from fasted sows using jugular venipuncture at weaning, and ELISA kits (Nanjing Jiancheng Bioengineering Institute, Nanjing, China) were used to determine serum LDH, IgG, SOD, HSP70, COR, and PRL concentrations. Sow milk samples (approximate 20 ml) were obtained on d 3, 15, and 20 of lactation from the last two pairs of sow nipples at 10:30–11:30am in each animal, and oxytocin was used to stimulate let-down. Three milk samples from each animal was pooled equally to evaluate the effect of environmental temperature on milk composition. In each environmental group, six animals that balanced for weaned backfat thickness and weight were chosen and humanely slaughtered at 21 day postpartum, and suckled mammary glands were split down the mid-line and tissues were excised from the center portion of four glands from the fourth and fifth pairs of nipples. Connective tissue and fat were removed. Mammary tissues were cut into small pieces and snap-frozen in liquid nitrogen prior to subsequent processing. In general, the collected mammary tissues contain primarily secretory epithelial cells, with a small amount of myoepithelial cells, endothelial cells, adipocytes, fibroblasts, and immune cells (Kensinger et al., 1986). All procedures were conducted under protocols approved by the Institutional Animal Care and Use Committee of South China Agricultural University, China.

RNA Preparation and Sequencing

Total RNA was extracted from mammary tissue and purified using Trizol reagent (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions. Each RNA sample was treated with DNase I (Takara, Dalian, China) for 15 minutes at 37°C to remove residual genomic DNA. RNA quantity and purity were analyzed using a Bioanalyzer 2100 (Agilent, Palo Alto, CA), and RNA samples with Integrity Number (RIN) value ≥ 7.5 were used for further analysis. In each experimental condition, we randomly selected two RNA samples and pooled 5 μg of RNA from each sample. In total, six RNA pools were depleted of ribosomal RNA using an Ribo-Zero™ rRNA Removal Kit (Illumina, San Diego, USA), and the left poly-$A^{-/+}$ RNA fractions were

then reverse-transcribed to create the final cDNA using a mRNA-Seq sample preparation kit (Illumina, San Diego, CA). Finally, we performed paired-end sequencing on an Illumina Hiseq 4000 (LC Bio, Hangzhou, China) to yield 2 × 150 nucleotide reads, following the manufacturer's recommended protocol.

Bioinformatics Analysis

Raw sequences quality was verified using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the reads that contained adaptor contamination, low quality, and undetermined bases were removed by Cutadapt (Martin, 2011). Filtered data from each library was aligned to the *Sscrofa11.1* reference genome downloaded from Ensembl genome website (ftp://ftp.ensembl.org/pub/release-94/fasta/sus_scrofa/dna/) with TopHat v2.1.1 (Kim et al., 2013), and transcript assembly and abundance estimation were performed using Cufflinks v2.2.1 (Trapnell et al., 2012). Each assembly was then merged using Cuffmerge to create a single transcriptome annotation with known porcine genes in gtf format (ftp://ftp.ensembl.org/pub/release-94/gtf/sus_scrofa) for subsequent protein-coding gene analysis. To predict circRNA candidates, five different algorithms including CIRCexplorer2 (Zhang X. et al., 2016), circRNA_Finder (Fu and Liu, 2014), CIRI (Gao et al., 2015), find_circ (Memczak et al., 2013), and MapSplice (Wang et al., 2010) were performed on each RNAseq library. Only circRNA candidates identified by all five approaches were considered for further evaluation. Following the above primary analysis, expression levels of all coding genes in each library were estimated from the TopHat alignments as fragments per kilobase of exon per million mapped reads (FPKM), and Cuffdiff, included in the Cufflinks package, was used to compare expression levels between NS and HS with a false discovery rate (FDR) value < 0.05. The abundance of circRNA candidates was calculated with back-spliced junction read count (Zhang et al., 2014). Then, the edgeR software package (Robinson et al., 2010) was used to examine the differential expression (DE) of circRNA candidates with P value < 0.05 and fold change ≥ 2. Finally, Biological Processes GO terms and KEGG pathway analysis of the DE genes were performed using DAVID gene functional classification (https://david.ncifcrf.gov/).

CeRNA Network Construction

Putative interactions between the DE circRNAs and lactation-related coding genes that responded to HS in our paper were evaluated by competing to bind with shared miRNAs. Porcine mature miRNAs published in miRBase (http://www.mirbase.org/) were prepared for further analysis. In details, the construction of ceRNA networks included three steps: (1) the correlations between DE circRNAs and lactation-related genes were calculated by the Pearson test, and only nodes in positive circRNA-gene interactions were retained; (2) the circRNA-miRNA and mRNA-miRNA interactions were predicted by miRanda algorithm (Betel et al., 2010) with with energies of ≤ −20.0 kcal/mol and no mismatch in the seed region (positions 2–8 in the 5′ end); (3) potential circRNA-miRNA-gene interactions were established and visualized using Cytoscape V3.4 (http://cytoscape.org/).

Validation of Sequencing Data by Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR)

Total RNA from the NS and HS samples were isolated with Trizol reagent (Invitrogen, Carlsbad, CA), and cDNA synthesis was conducted using PrimeScript RT reagent Kit (Takara, Dalian, China) with random hexamers. Quantitative PCR was used to analyze the expression changes of the chosen transcripts with SYBR Premix Ex Taq II (Takara). All primers are listed in **Table S1**, and final expression data were calculated using the $2^{-\Delta Ct}$ method using porcine GAPDH as a reference gene.

Sequencing Data Submission

All sequencing raw datasets have been deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/) with the BioProject accession number PRJNA578241.

Statistical Analysis

The statistical analysis was performed using by SPSS 17.0 Statistics Software (Chicago, IL, USA). The results of ELISA and RT-qPCR analysis between two groups were compared with independent t-test; the correlation analyses of DE circRNAs with lactation-related coding genes were tested by function cor (x, y, use = "p"), and illustrated with function labeledHeatmap (Matrix, xLabels, yLabels) in R package WGCNA (http://127.0.0.1:11153/library/stats/html/cor.html).

RESULTS

Sow and Litter Performance

This study was performed during December 2016 and during August 2017. During the winter experimental stage, ambient temperatures and relative humidity averaged 14.3 ± 0.81°C and 65.0% ± 0.69%, while the corresponding values for the hot season were 32.7 ± 1.40°C and 76.1% ± 0.38%, respectively. The effects of ambient temperature on the serum stress-associated variables of lactating sows are presented in **Figure S1**. Blood heat shock protein 70 (HSP70), lactate dehydrogenase (LDH), and immunoglobulin G (IgG) levels were significantly higher in the HS cohort compared with the NS group (P < 0.05). In contrast, serum superoxide dismutase (SOD) and prolactin (PRL) concentrations were lower in the HS population (P < 0.05). The high temperature group had significantly lower serum cortisol (COR) concentration than the NS group (P < 0.05). In addition, there was no effect of thermal stress on the number of piglets born alive, litter birth weights, or piglets alive per litter at weaning (P > 0.05, **Table 1**). In contrast, litter weights at weaning were significantly higher in the NS cohort than the HS cohort (P < 0.05). And a reduced average daily weight gain of piglets (193.9 ± 2.19 vs. 218.0 ± 1.89 g, respectively, for the HS and NS cohorts) was associated with the environment with a high temperature. Thermal stress also altered the milk composition of lactating sows. In particular, sow milk had less butterfat, and lactose when sows lactated in the hot environment (P < 0.05), and milk in the

TABLE 1 | Production traits of tested sows and piglets between summer and winter.

Variables	Summer (N = 30)	Winter (N = 30)
Number born alive	10.8 ± 0.13	10.9 ± 0.09
Weaning survival	10.4 ± 0.13	10.5 ± 0.11
Litter birth weight, kg	14.9 ± 0.18	15.2 ± 0.17
Weight of weaning litter, kg	61.3 ± 0.73^{B}	67.2 ± 1.59^{A}
Average daily gain, g	193.9 ± 2.19^{B}	218.0 ± 1.89^{A}

A and B denote values that differ significantly at $P < 0.01$.

HS group had a downward trend for total milk solids and milk protein levels (**Figure 1**). HS individuals tended to have much lower casein alpha s1 (CSNαs1) and s2 (CSNαs2) distributions ($P < 0.01$), while casein beta (CSNβ) and casein kappa (CSNκ) concentrations decreased from 343.59 ± 6.42 μg/ml and 7.47 ± 0.16 μg/ml in the HS group to 259.14 ± 7.96 μg/ml and 6.35 ± 0.12 μg/ml in the NS group ($P < 0.01$), respectively. Under HS, we found no overall differences in whey acidic protein (WAP) concentrations between the NS and HS groups ($P = 0.067$), although there was a slightly decreasing trend in the HS group (**Table 2**).

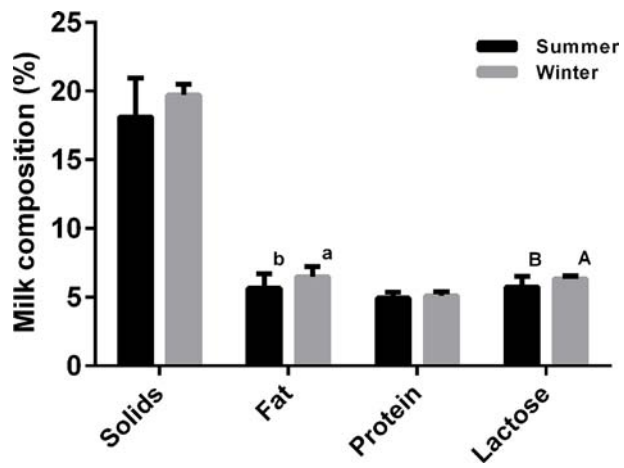


FIGURE 1 | Effect of thermal stress on milk composition of lactating sows. a and b denote values that differ significantly at $P < 0.05$, and A and B denote values that differ significantly at $P < 0.01$ (N = 30).

TABLE 2 | Effect of thermal stress on lactoprotein distribution of lactating sows.

Variables	Summer (N = 30)	Winter (N = 30)
CSNαs1, μg/ml	592.03 ± 23.31^{B}	693.63 ± 16.35^{A}
CSNαs2, μg/ml	282.31 ± 15.00^{B}	409.85 ± 9.27^{A}
CSNβ, μg/ml	259.14 ± 7.96^{B}	343.59 ± 6.42^{A}
CSNκ, μg/ml	6.35 ± 0.12^{b}	7.47 ± 0.16^{a}
WAP, ng/ml	447.87 ± 14.67	484.14 ± 13.65

CSNαs1, casein alpha s1; CSNαs2, casein alpha s2; CSN2, casein beta; CSN3, casein kappa; WAP, whey acidic protein. a and b denote values that differ significantly at $P < 0.05$, and A and B denote values that differ significantly at $P < 0.01$.

RNA Sequencing and Transcript Analysis

Six cDNA libraries were constructed from porcine mammary tissues exposed to thermal stress or to suitable temperatures. Each RNA-seq library generated approximately 94.52 ± 2.83 million raw reads of 110 nt in length, representing about 3.70 ± 0.11 fold coverage of the porcine genome. After quality control trimming, a total of 88.08 ± 2.65 million valid reads were obtained, accounting for 93.19% ± 0.23% of the raw reads in each library. We aligned all valid reads onto the porcine *Sscrofa11.1* reference genome and found that over 80.55% ± 1.63% of the reads could be mapped successfully to the genome, including 76.13% ± 1.63% of the mapped reads with proper pair alignment in the six libraries (**Table S2A**). In addition, most valid reads were mapped to exons (77.51% ± 2.64%), 17.27% ± 2.02% were mapped to introns, and the rest to intergenic regions (5.22% ± 0.62%), indicating confidence in the quality of library construction and sequencing analysis.

Transcript assemblies from porcine mammary tissue with Cufflinks revealed a total of 133,145 isoforms across six samples, including approximately 37.33% identified candidates that completely matched Ensembl transcript regions (**Table S2B**). A comparison of known Ensembl transcripts revealed that 36,271 isoforms were expressed across all tissues; NS-specific units accounted for 82.08% of known Ensembl isoforms, while 81.37% of the known isoforms existed in the HS libraries (**Table S2C**). Raw Ensembl gene expression levels were quantified by the FPKM algorithm, and the 10 most prevalent functional isoforms accounted for 7.97% ± 0.62% of the total raw reads. In addition, seven gene candidates of *PAEP*, *CSN1S1*, *CSN3*, *CSN2*, *JCHAIN*, *COX1,* and *NUPR1* were shared in the top 10 expressed genes in each sequencing library. These highly expressed isoforms are well-known as having key functions in the lactation process, consistent with the physiological roles of genes expected to be found in mammary gland tissues.

Identification of circRNA

Several tools have been developed for identification of circRNAs based on back-spliced reads produced from high-throughput RNA sequencing datasets (Hansen et al., 2015). Due to the rearranged exon ordering, these back-spliced events usually cannot be mapped onto the reference genomes (Zhang et al., 2014). In the present study, we identified 17.07 ± 0.25 (19.75% ± 0.80% of the valid reads) and 17.07 ± 3.19 (19.16% ± 3.55%) million unmapped reads in the NS and HS libraries, respectively. Among them, we found a total of 948.00 ± 23.98 thousand back-spliced junction events (1.09% ± 0.04% of the valid reads) in the rRNA-depleted libraries of NS animals, as well as 892.49 ± 80.27 thousand candidates in the HS libraries. We then compared five different circRNA predicting algorithms and found a total of 31,031 unique circRNAs identified across six libraries. Of these, 19,642 were found by a single algorithm, accounting for 63.29% of all the circRNAs identified (**Figure 2**), indicating that the circRNA landscape differs quite dramatically depending on the algorithm used. In particular, find_circ and MapSplice exhibited the highest and lowest level of sensitivity; this is in part reflected in the total number of circRNAs predicted, where find_circ and

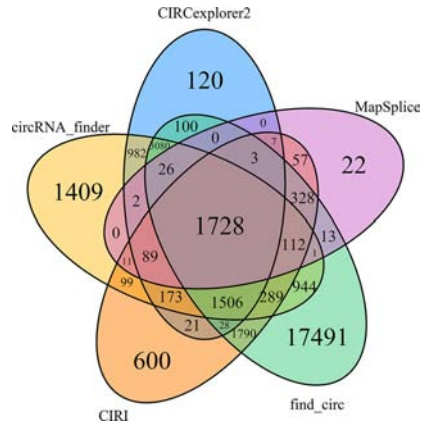


FIGURE 2 | Common circRNA candidates identified by all five approaches.

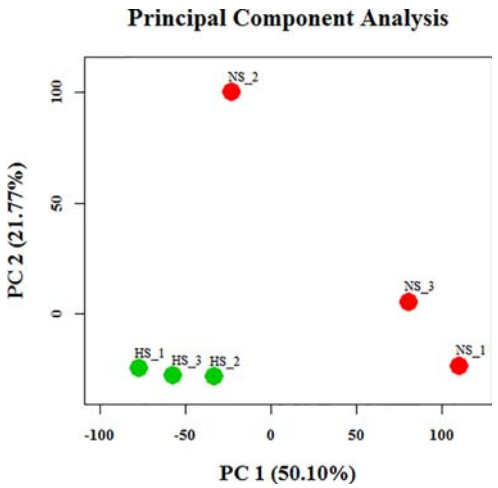


FIGURE 3 | Principal component analysis of assembled transcripts in six libraries. PC, principal component; NS, non-heat stress; HS, heat stress.

MapSplice output the highest and lowest number of circRNA species (27,439 and 2,399, respectively) compared to CIRCexplorer2, circRNA_finder, and CIRI methods (7,865, 10,451, and 6,841 species, respectively; **Table S3**). To limit the level of false positive circRNAs, only circRNA candidates identified by all five approaches were considered for further evaluation. Of the 31,031 predicted circRNAs, only a modest overlap of 1,728 circRNAs (5.57%) was observed among the five prediction pipelines. These 1,728 circularization events were found to be produced from 1,157 hosting isoform loci, including 571 transcripts that generated more than one circRNA. For instance, we found that the porcine genes *SEC24A* and *SLC5A10* had nine and eight predicted circRNAs, respectively, and there were seven circularization events predicted from the *BAZ2B*, *PIAS1,* and *CCAR1* genes (**Table S4A**).

Differential Expression and Functional Analysis

To identify dissimilarities between the tested individuals, principal component analysis (PCA) of the globally expressed transcript with the FPKM levels was performed (**Figure 3**). This analysis indicated that the differences in expression between the NS and HS groups were greater than the differences between pools from each particular group. Therefore, we employed the Cuffdiff algorithm to analyze differences in mammary gland gene expression between the NS and HS groups to identify candidate transcripts that are responsive to thermal stress. In our dataset there were over 40,000 unique Ensembl transcripts sequenced, most of which had a very small FPKM value in total across all libraries. In order to filter out false-positive results, we only kept confident transcripts that were expressed in at least three libraries, and finally 9,789 tags were identified in our study (**Table S5A**). Among these, we detected a total of 67 differentially expressed (DE) transcript events by a limited cut-off of FDR < 0.05, representing 63 protein-coding genes, with 42 transcripts increased and 25 transcripts decreased in the HS groups compared to the NS group (**Table S5A**). Analysis of gene ontology (GO) enrichment for DE genes, using identified porcine genes as background in the current experiment, revealed that these genes were significantly enriched in lactation-related functions or stress-inducible biological processes, including "lactation", "defense response", "inflammatory response", "response to stimulus", and "regulation of immune system process" (**Table S5B**). These 67 DE genes are significantly involved in only one KEGG signaling pathways, termed as toll-like receptor signaling pathway. Although the role of these genes needs to be validated experimentally, the GO and KEGG pathway analyses collectively illustrate some possible avenues to improve HS resistance of lactating sows. In addition, we also clustered differential porcine circRNA expression counts between NS and HS libraries, as determined by the CIRCexplorer2 algorithm, and normalized with trimmed mean of M-values (TMM) (Robinson et al., 2010). In total, only 50 out of 1,728 identified circRNAs were DE, including 29 up-regulated candidates and 21 down-regulated candidates in the NS samples compared with the HS samples (**Table S4B**).

Functional Interactions Between circRNA and mRNA

To identify possible correlations between circRNA and mRNA expression, we first used the Pearson test and found a total of 1,423 significant interactions between DE circRNAs and DE genes in our study **(Table S6**). In **Table S6** we identified 464 sponge modulators participating in 100 miRNA-mediated regulatory interactions, including 45 circRNAs and 36 unique mRNAs. In addition, the Pearson analysis also revealed 84 significant interactions between DE circRNAs and four lactation-related coding genes (CSN1S1, CSN1S2, CSN3, WAP) that were annotated by GO analyses, and these interactions included 42 positive significant interactions and 42 negative interactions (**Table S6**). We observed that the highest expressed circRNA, *circCSN1S1_2*, was significantly and

positively associated with the expression of the *CSN1S1*, *CSN1S2*, *CSN3,* and *WAP* genes; interestingly, these gene products represent the main components of lactoprotein. Recent reports also showed that diverse RNA species can communicate with and co-regulate each other by competing to bind with shared miRNAs, acting as competing endogenous RNAs (ceRNAs) (Tay et al., 2014). We constructed circRNA-miRNA-mRNA networks by pairing the shared miRNA recognition sequences. In total, we generated a network that contained eight nodes and five connections formed between four circRNAs, three miRNAs, and *CSN1S1* gene, including five circRNA-miRNA interactions and three mRNA-miRNA interactions (**Table S7**). Of these, the circCSN1S1_2-miR-204 (miR-670)-CSN1S1 ceRNA axis was established, corresponding well with the ceRNA hypothesis.

RT-qPCR Analysis

Based on ceRNA networks involved in mammopoiesis and lactogenesis under HS that we constructed, we identified a total of 10 interaction core genes (**Figure S2**). We validated expression of these core genes by RT-qPCR, including eight lactation-related coding genes, one heat-response gene, and circCSN1S1_2, which had the highest expression levels in our study. Divergent and convergent primers were designed for circRNA candidates according to the method described in previous study (Sun et al., 2017). All tested candidate genes showed consistent expression patterns between RT-qPCR and sequencing analysis, suggesting that our estimation of abundance was accurate. Briefly, the expression of coding genes for lactoprotein were significantly lower in HS sows ($P < 0.05$), except for *CSN3*, which decreased in HS sows, but the change was not significant. Usually, HSP family member proteins have important roles as molecular chaperones that help prevent apoptosis under various stress conditions, including HS (Sakatani et al., 2012). In agreement with the sequencing analysis, HSP90AA1 showed a strong induction in response to HS, and it had high expression levels in the HS group. In addition, expression of circCSN1S1_2 was significantly lower in HS sows, demonstrating the validity of our post-transcriptional ceRNA regulatory model.

DISCUSSION

In general, HS is caused by a combination of environmental temperature, relative humidity, solar radiation, air movement and precipitation, and the majority of studies on HS in livestock have focused mainly on temperature and relative humidity (Bohmanova et al., 2007). In our experiment, the average temperatures and humidity levels during the HS challenge were 32.7 ± 0.40°C and 76.1 ± 0.38%, respectively; thus, the lactating sows in our test might be under HS (Bergsma and Hermesch, 2012). We observed greater concentrations of LDH and HSP70, and a decrease of COR levels in sows under HS conditions, compared with those in an NS climate; these values are generally considered as indicators of stress in pigs (Yu et al., 2010; Belhadj Slimen et al., 2016). Enhanced levels of serum LDH is a biomarker of liver damage in hyperthermic animals (Ozaki et al., 1995). Cao et al. (2011) reported that chronic HS induces significant increases LDH levels in rat plasma; this report was similar to what we observed in lactating sows. HSPs are molecular chaperones that differ in regards to their biological functions and molecular weights (Feder and Hofmann, 1999). Among the various HSP classes, HSP70 levels are associated with the acquisition of thermotolerance (Bedulina et al., 2013). In farm animals, HS significantly increases serum HSP70 concentrations in beef cattle (Gaughan et al., 2013), dairy cow (Min et al., 2015), buffalo (Manjari et al., 2015), sheep (Romero et al., 2013), goat (Dangi et al., 2015), and broiler chicken (Gu et al., 2012); these reports are in excellent agreement with the experimental results in our study. In addition, Wang et al. (2015) reported that acute heat exposure significantly elevated levels of COR in rats, and porcine serum COR levels rapidly increased when individuals were subjected to 40°C for 5 hours (Yu et al., 2010). In contrast, summer temperature-induced HS dramatically repressed COR concentrations in the present study. This finding may be due to the different effects between short-term acute HS and chronic HS. Generally, short-term heat exposure increases plasma COR levels, while long-term exposure decreases them (Du Preez, 2000). Christison and Johnson (1972), and Wiersma and Stott (1974) noted a similar trend in dairy cattle exposed to hot summer conditions. HS has been suggested to be responsible for inducing oxidative stress and immune responses in livestock animals during the summer (Das et al., 2016). In the present study, serum SOD and IgG levels were higher in the HS group than in the NS group, suggesting that the antioxidative and immune function of sows increases to adapt to the adverse environment. Recently, elevated concentrations of SOD and IgG were also reported to be sensitive to ambient temperature in broiler chickens (Azad et al., 2010), lactating buffaloes, and cows (Lallawmkimi et al., 2013; Yatoo et al., 2014). PRL is vital for lactogenesis (Akersr et al., 1981), and concentrations of plasma PRL decrease in dairy cows during thermal stress (Tao et al., 2011). In agreement with previous studies, our data demonstrated a significant reduction in porcine PRL levels due to elevated ambient temperature.

In our study, no significant effects of HS were observed on the number of live piglets born per litter, nor on litter birth weight; similar findings were reported by Lucy and Safranski (2017), who demonstrated no clear influences of gestational HS on the number of piglets born live per litter. The seasonal influences of our study and that have been reported previously on piglet traits at birth are mainly caused by a delayed response to ambient temperature, i.e. sows were mated and conceived during the cool season and subsequently farrowed in the hot season; it is established that the primary effects of temperature on litter traits in piglets primarily occur during the first 4 weeks of gestation (Tummaruk et al., 2004). The HS sows weaned piglets that were approximately 0.5 kg lighter in our study than the NS sows. This represented an about 8.47% decrease in weaning weight, and is in accordance with the results of Williams et al. (2013). In addition, piglets were more susceptible to heat-induced reductions in piglet weight gain during early lactation, in concordance with a study by Spencer

et al. (2003), who reported a 17% decrease in piglet weaning weight when lactation period lasted 14 days. We observed a downward trend of high temperature on total milk solids, which were reduced by approximately 8.77%, including specific losses in milk fat and lactose content by about 14.69% and 10.45%, respectively. Generally, milk composition varies considerably throughout the seasons, as showed in multiple farm animals included Holstein cows (Shwartz et al., 2009; Bernabucci et al., 2015), dairy goats (Chornobai et al., 1999), and mares (Markiewicz-Kęszycka et al., 2015). In dairy cows, lower milk fat (Bernabucci et al., 2015) and lactose (Shwartz et al., 2009) values are recorded during the summer months, in correspondence with an increase in Temperature Humidity Index. A similar trend of variation in milk composition was also reported for goat milk, and high air temperature in the summer was significantly negatively correlated with goat milk physico-chemical characteristics (Chornobai et al., 1999). In particular, the milk characteristics most highly affected by air temperature were fat and lactose contents, with correlation coefficients of −0.90 and −0.77, respectively. In contrast, mare milk collected in summer had a significantly higher fat content than in autumn, but the average concentration of lactose was similar for milk collected in summer and in autumn, and showed no specific significant seasonal variations (Markiewicz-Kęszycka et al., 2015). These large discrepancies observed in lactating mares may be due to differences in experimental animals (sows and cows vs. mares) or differences in climate conditions (32.7 ± 0.40°C vs. 23.15 ± 1.61°C in the summer). In agreement with a study comparing lactating sows exposed to high ambient temperature (Renaudeau and Noblet, 2001), our results showed no significant effects of elevated ambient temperature on milk protein, but HS during summer significantly decreased CSNαs1, CSNαs2, and CSNκ concentrations in milk. These results confirmed those obtained by Bernabucci et al. (2015) carried out in dairy cows, in which it was reported that the concentration of CSNα during summer months was 22.6% lower than in the winter, and was 16% lower than in the spring. CSNκ levels were also 9.7% lower during summer than in the winter. Our study agrees with previous studies, and indicates that there is a significant seasonal effect on CSN fractions in domestic livestock milk.

Generally, heat-stressed lactating sows reduce their feed intake, leading to loss of milk production, which can negatively affect piglet growth and development during lactation (Ribeiro et al., 2018). However, Rhoads et al. (2009) and Wheelock et al. (2010) have recently demonstrated that reduced nutrient intake only accounts for about 35%–50% of the HS-induced decrease in milk synthesis. A large portion of the thermal effects on animal lactation may be a consequence of energy intake-independent changes (Wheelock et al., 2010), resulting from genetic regulation of nutrient partitioning during HS (Collier et al., 2008). In the current study, we therefore used RNA-Seq to find the underlying molecular mechanisms of milk synthesis under HS in lactating sows. Sequencing revealed a total of 19,032 unique Ensembl genes in lactating porcine mammary tissues, while genes encoding caseins, whey proteins, and enzymes involved in lactogenesis pathways showed higher expression than other genes with RPKM values. Similar results were obtained in cows (Wickramasinghe et al., 2012), goats (Shi et al., 2015; Crisà et al., 2016), and humans (Lemay et al., 2013). A total of 16,892 genes were expressed in bovine milk somatic cells during early lactation, as well as 19,094 in peak lactation and 18,070 in late lactation, and *LGB* (β-lactoglobulin), *CSN2* (β-casein), *CSN1S1* (α-S1-casein), *LALBA* (α-lactalbumin), *CSN3* (κ-casein), and *CSN1S2* (casein-α-S2) were identified as having the highest expression in milk, based on RPKM values (Wickramasinghe et al., 2012). Approximately 16,024 ovine NCBI unigenes were found to be expressed in mammary glands (Shi et al., 2015), and *CSN2*, *CSN3*, *CSN1S1*, *CSN1S2*, *LALBA,* and *LGB* were the most abundant in the mammary gland transcriptome (Shi et al., 2015; Crisà et al., 2016). In human mammary cells, Lemay et al. (2013) reported 14,629, 14,529, and 13,745 unique genes expressed in colostral, transitional, and mature stages of lactation, and β-casein and α-lactalbumin transcripts made up 45% of the total mRNA expression during lactation. Of the top genes identified in our study, *CSN1S2* had the highest expression among the *CSN* family, followed by *CSN2*, *CSN1S1*, and *CSN3*, which were in discordance with the composition of caseins identified by ELISA analysis. We found that porcine casein-α-S1 constituted up to 51.94% of the caseins in our study. One possibility for this discrepancy was that the abundant caseins are broken into bio-active peptides, and therefore their concentrations are not accurately reflected in the analysis of major milk component proteins. The expression of bio-active peptides formed by cleavage of caseins are higher toward the beginning of lactation (Silva and Malcata, 2005). Another possible explanation is that even though there was high expression of the genes encoding caseins, the protein synthesis may not be efficient in sows that were in negative energy balance, or that were limited in dietary intake of essential amino acids (Wickramasinghe et al., 2012).

In order to further reveal the mechanism of response of lactating sows to HS, we focused on identification of differently expressed genes in response to high ambient temperature. Functional annotation analysis identified that four of these DE genes have principal roles in lactogenesis, including four down-regulated genes (*CSN1S1*, *CSN1S2*, *CSN3,* and *WAP*) in the heat stressed individuals. The CSN1S1, CSN1S2, CSN3, and WAP proteins are the main components of lactoprotein that is usually reduced in response to HS in dairy animals (Bernabucci et al., 2015), and the gene expression analysis was in accordance with the results of the ELISA assay.

Recently, a class of non-coding RNAs, called circRNAs, has been identified across the animal kingdoms (Memczak et al., 2013). These circRNAs usually act as ceRNAs to regulate other coding genes by sharing specific miRNA binding sites (Salmena et al., 2011). Multiple types of circRNA-mediated ceRNA interactions have been linked to various physiological or pathological states, including members of the miRNA-2284 family that are sponged by circCSN1S1 to regulate bovine casein translation (Zhang et al., 2015). Therefore, in the

present study we carried out a circRNAome analysis of porcine mammary tissues between NS and HS groups. We identified 50 candidate circRNAs out of 1,728 identified circRNAs that were DE between the NS and HS groups. Based on the ceRNA hypothesis, 42 positively correlated circRNA-mRNA interactions were constructed between the four lactogenic genes and the DE circRNAs using the Pearson algorithm. Of these interaction pairs, analysis by the miRanda application (Enright et al., 2003) revealed four circRNA-mRNA interactions that were predicted to share the same miRNA regulatory elements. In particular, the circCSN1S1_2 binds competitively with miR-204 to increase expression of its hosting gene, *CSN1S1*. A similar ceRNA network was strongly suggested between circFoxo3 and *Foxo3* mRNA in tumor growth and angiogenesis (Yang et al., 2016). Yang et al. reported that circFoxo3 shared identical sequences with the *Foxo3* gene to bind miR-22, miR-136, miR-138, miR-149, miR-433, miR-762, miR-3614-5p, and miR-3622b-5p. These observations indicated that the expression of circRNAs might be related to the expression of their parental genes. Taken together, several circRNA-miRNA-mRNA axes were shown to be likely involved in porcine lactogenesis under HS, and these findings provide novel perspectives on circRNA-associated ceRNA networks for future research in sow lactation.

CONCLUSION

We found that constant elevated ambient temperature and HS has negative consequences on piglet growth and performances due to decreased milk production and characteristics of lactating sows. Thermal stress altered genome-wide profiles of circRNAs dramatically in lactating porcine mammary tissue, and these heat-induced circRNAs might participate in mammopoiesis and lactogenesis by post-transcriptional regulation of ceRNA networks. Our results provide novel rationale to investigate circRNA functions in the lactating sow response to HS, and additional research is necessary to quantify and understand these effects.

DATA AVAILABILITY STATEMENT

All sequencing raw datasets have been deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/) with the BioProject accession number PRJNA578241.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of South China Agricultural University, China.

AUTHOR CONTRIBUTIONS

All authors were involved in project conception and design. JS led the lab assays, analyses of data, and writing of the manuscript. JS and YZ contributed reagents, materials, and analysis tools. All authors gave final approval for publication.

FUNDING

This research was financially supported by the National Key Research and Development Program of China (2016YFD0500503), the Natural Science Foundation of China Program (31802032), the major scientific projects in general colleges and Universities of Guangdong Province (2017KTSCX023), and the Natural Science Foundation of Guangdong Province (2018B030311015).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01347/full#supplementary-material

REFERENCES

Akersr, M. R., Bauman, D. E., Capuco, A. V., Goodman, G. T., and Tucker, A. H. (1981). Prolactin regulation of milk secretion and biochemical differentiation of mammary epithelial cells in periparturient cows. *Endocrinology* 109 (1), 23–30. doi: 10.1210/endo-109-1-23

Azad, M. A. K., Kikusato, M., Maekawa, T., Shirakawa, H., and Toyomizu, M. (2010). Metabolic characteristics and oxidative damage to skeletal muscle in broiler chickens exposed to chronic heat stress. *Comp. Biochem. Physiol. Part A: Mol. Integr. Physiol.* 155 (3), 401–406. doi: 10.1016/j.cbpa.2009.12.011

Bedulina, D. S., Evgen'Ev, M. B., Timofeyev, M. A., Protopopova, M. V., Garbuz, D. G., Pavlichenko, V. V., et al. (2013). Expression patterns and organization of the hsp70 genes correlate with thermotolerance in two congener endemic amphipod species (Eulimnogammarus cyaneus and E. verrucosus) from Lake Baikal. *Mol. Ecol.* 22 (5), 1416–1430. doi: 10.1111/mec.12136

Belhadj Slimen, I., Najar, T., Ghram, A., and Abdrrabba, M. (2016). Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. *J. Anim. Physiol. Anim. Nutr.* 100 (3), 401–412. doi: 10.1111/jpn.12379

Bergsma, R., and Hermesch, S. (2012). Exploring breeding opportunities for reduced thermal sensitivity of feed intake in the lactating sow. *J. Anim. Sci.* 90 (1), 85–98. doi: 10.2527/jas.2011-4021

Bernabucci, U., Basiricò, L., Morera, P., Dipasquale, D., Vitali, A., Cappelli, F. P., et al. (2015). Effect of summer season on milk protein fractions in Holstein cows. *J. Dairy Sci.* 98 (3), 1815–1827. doi: 10.3168/jds.2014-8788

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11 (8), R90. doi: 10.1186/gb-2010-11-8-r90

Bohmanova, J., Misztal, I., and Cole, J. B. (2007). Temperature-humidity indices as indicators of milk production losses due to heat stress. *J. Dairy Sci.* 90 (4), 1947–1956. doi: 10.3168/jds.2006-513

Cao, D., Li, H., Yi, J., Zhang, J., Che, H., Cao, J., et al. (2011). Antioxidant properties of the mung bean flavonoids on alleviating heat stress. *PloS One* 6 (6), e21071. doi: 10.1371/journal.pone.0021071

Chornobai, C. A., Damasceno, J. C., Visentainer, J. V., and Matsushita, M. (1999). Physical-chemical composition of in natura goat milk from cross Saanen throughout lactation period. *Archiv. Latinoamericanos Nutricion* 49 (3), 283–286.

Christison, G. I., and Johnson, H. D. (1972). Cortisol turnover in heat-stressed cows. *J. Anim. Sci.* 35 (5), 1005–1010. doi: 10.2527/jas1972.3551005x

Collier, R. J., Dahl, G. E., and VanBaale, M. J. (2006). Major advances associated with environmental effects on dairy cattle. *J. Dairy Sci.* 89 (4), 1244–1253. doi: 10.3168/jds.S0022-0302(06)72193-2

Collier, R. J., Collier, J. L., Rhoads, R. P., and Baumgard, L. H. (2008). Invited review: genes involved in the bovine heat stress response. *J. Dairy Sci.* 91 (2), 445–454. doi: 10.3168/jds.2007-0540

Crisà, A., Ferrè, F., Chillemi, G., and Moioli, B. (2016). RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC Veterinary Res.* 12 (1), 264. doi: 10.1186/s12917-016-0881-7

Dangi, S. S., Gupta, M., Dangi, S. K., Chouhan, V. S., Maurya, V. P., Kumar, P., et al. (2015). Expression of HSPs: an adaptive mechanism during long-term heat stress in goats (Capra hircus). *Int. J. Biometeorol.* 59 (8), 1095–1106. doi: 10.1007/s00484-014-0922-5

Das, R., Sailo, L., Verma, N., Bharti, P., and Saikia, J. (2016). Impact of heat stress on health and performance of dairy animals: a review. *Veterinary World* 9 (3), 260. doi: 10.14202/vetworld.2016.260-268

Du Preez, J. H. (2000). Parameters for the determination and evaluation of heat stress in dairy cattle in South Africa. *Onderstepoort J. Vet Res.* 67 (4), 263–271.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. *Genome Biol.* 5 (1), R1. doi: 10.1186/gb-2003-5-1-r1

Feder, M. E., and Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* 61 (1), 243–282. doi: 10.1146/annurev.physiol.61.1.243

Fu, X., and Liu, R. (2014). "CircRNAFinder: a tool for identifying circular RNAs using RNA-Seq data" in *Proceedings of the 6th International Conference on Bioinformatics and Computational Biology.* (Las Vegas, Nevada, USA: BICOB).

Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for *de novo* circular RNA identification. *Genome Biol.* 16 (1), 4. doi: 10.1186/s13059-014-0571-3

Gaughan, J. B., Bonner, S. L., Loxton, I., and Mader, T. L. (2013). Effects of chronic heat stress on plasma concentration of secreted heat shock protein 70 in growing feedlot cattle. *J. Anim. Sci.* 91 (1), 120–129. doi: 10.2527/jas.2012-5294

Gu, X. H., Hao, Y., and Wang, X. L. (2012). Overexpression of heat shock protein 70 and its relationship to intestine under acute heat stress in broilers: 2. Intestinal oxidative stress. *Poultry Sci.* 91 (4), 790–799. doi: 10.3382/ps.2011-01628

Hanan, M., Soreq, H., and Kadener, S. (2017). CircRNAs in the brain. *RNA Biol.* 14 (8), 1028–1034. doi: 10.1080/15476286.2016.1255398

Hansen, T. B., Venø, M. T., Damgaard, C. K., and Kjems, J. (2015). Comparison of circular RNA prediction tools. *Nucleic Acids Res.* 44 (6), e58–e58. doi: 10.1093/nar/gkv1458

Hill, D. L., and Wall, E. (2015). Dairy cattle in a temperate climate: the effects of weather on milk yield and composition depend on management. *Animal* 9 (1), 138–149. doi: 10.1017/S1751731114002456

Kadzere, C. T., Murphy, M. R., Silanikove, N., and Maltz, E. (2002). Heat stress in lactating dairy cows: a review. *Livestock Prod. Sci.* 77 (1), 59–91. doi: 10.1016/S0301-6226(01)00330-X

Kensinger, R. S., Collier, R. J., and Bazer, F. W. (1986). Ultrastructural changes in porcine mammary tissue during lactogenesis. *J. Anat.* 145, 49.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14 (4), R36. doi: 10.1186/gb-2013-14-4-r36

Lai, X., Bazin, J., Webb, S., Crespi, M., Zubieta, C., and Conn, S. J. (2018). "CircRNAs in plants," in *Circular RNAs* (Singapore: Springer), 329–343.

Lallawmkimi, M. C., Singh, S. V., Upadhyay, R. C., and De, S. (2013). Impact of vitamin E supplementation on heat shock protein 72 and antioxidant enzymes in different stages of Murrah buffaloes during seasonal stress. *Indian J. Anim. Sci.* 83 (9), 909–915.

Lemay, D. G., Ballard, O. A., Hughes, M. A., Morrow, A. L., Horseman, N. D., and Nommsen-Rivers, L. A. (2013). RNA sequencing of the human milk fat layer transcriptome reveals distinct gene expression profiles at three stages of lactation. *PloS One* 8 (7), e67531. doi: 10.1371/journal.pone.0067531

Lucy, M. C., and Safranski, T. J. (2017). Heat stress in pregnant sows: thermal responses and subsequent performance of sows and their offspring. *Mol. Reprod. Dev.* 84 (9), 946–956. doi: 10.1002/mrd.22844

Manjari, R., Yadav, M., Ramesh, K., Uniyal, S., Rastogi, S. K., Sejian, V., et al. (2015). HSP70 as a marker of heat and humidity stress in Tarai buffalo. *Trop. Anim. Health Prod.* 47 (1), 111–116. doi: 10.1007/s11250-014-0692-4

Markiewicz-Kęszycka, M., Czyżak-Runowska, G., Wójtowski, J., Jóźwik, A., Pankiewicz, R., Łęska, B., et al. (2015). Influence of stage of lactation and year season on composition of mares' colostrum and milk and method and time of storage on vitamin C content in mares' milk. *J. Sci. Food Agric.* 95 (11), 2279–2286. doi: 10.1002/jsfa.6947

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17 (1), 10–12. doi: 10.14806/ej.17.1.200

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495 (7441), 333. doi: 10.1038/nature11928

Messias de Bragança, M., Mounier, A. M., and Prunier, A. (1998). Does feed restriction mimic the effects of increased ambient temperature in lactating sows. *J. Anim. Sci.* 76 (8), 2017–2024. doi: 10.2527/1998.7682017x

Min, L., Cheng, J. B., Shi, B. L., Yang, H. J., Zheng, N., and Wang, J. Q. (2015). Effects of heat stress on serum insulin, adipokines, AMP-activated protein kinase, and heat shock signal molecules in dairy cows. *J. Zhejiang University-Sci. B* 16 (6), 541–548. doi: 10.1631/jzus.B1400341

Ozaki, M., Fuchinoue, S., Teraoka, S., and Ota, K. (1995). The *in vivo* cytoprotection of ascorbic acid against ischemia/reoxygenation injury of rat liver. *Arch. Biochem. Biophysics* 318 (2), 439–445. doi: 10.1006/abbi.1995.1252

Pan, T., Sun, X., Liu, Y., Li, H., Deng, G., Lin, H., et al. (2018). Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis. *Plant Mol. Biol.* 96 (3), 217–229. doi: 10.1007/s11103-017-0684-7

Patop, I. L., and Kadener, S. (2018). circRNAs in cancer. *Curr. Opin. In Genet. Dev.* 48, 121–127. doi: 10.1016/j.canlet.2018.03.035

Quiniou, N., and Noblet, J. (1999). Influence of high ambient temperatures on performance of multiparous lactating sows. *J. Anim. Sci.* 77 (8), 2124–2134. doi: 10.2527/1999.7782124x

Renaudeau, D., and Noblet, J. (2001). Effects of exposure to high ambient temperature and dietary protein level on sow milk production and performance of piglets. *J. Anim. Sci.* 79 (6), 1540–1548. doi: 10.2527/2001.7961540x

Renaudeau, D., Quiniou, N., and Noblet, J. (2001). Effects of exposure to high ambient temperature and dietary protein level on performance of multiparous lactating sows. *J. Anim. Sci.* 79 (5), 1240–1249. doi: 10.2527/2001.7951240x

Renaudeau, D., Collin, A., Yahav, S., De Basilio, V., Gourdine, J. L., and Collier, R. J. (2012). Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Animal* 6 (5), 707–728. doi: 10.1017/S1751731111002448

Renaudeau, D. (2005). Effects of short-term exposure to high ambient temperature and relative humidity on thermoregulatory responses of European (Large White) and Caribbean (Creole) restrictively-fed growing pigs. *Anim. Res.* 54 (2), 81–93. doi: 10.1051/animres:2005005

Rhoads, M. L., Rhoads, R. P., VanBaale, M. J., Collier, R. J., Sanders, S. R., Weber, W. J., et al. (2009). Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism, and aspects of circulating somatotropin. *J. Dairy Sci.* 92 (5), 1986–1997. doi: 10.3168/jds.2008-1641

Ribeiro, B. P. V. B., Lanferdini, E., Palencia, J. Y. P., Lemes, M. A. G., de Abreu, M. L. T., de Souza Cantarelli, V., et al. (2018). Heat negatively affects lactating swine: a meta-analysis. *J. Thermal Biol.* 74, 325–330. doi: 10.1016/j.jtherbio.2018.04.015

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Romero, R. D., Pardo, A. M., Montaldo, H. H., Rodríguez, A. D., and Cerón, J. H. (2013). Differences in body temperature, cell viability, and HSP-70 concentrations between Pelibuey and Suffolk sheep under heat stress. *Trop. Anim. Health Prod.* 45 (8), 1691–1696. doi: 10.1007/s11250-013-0416-1

Sakatani, M., Alvarez, N. V., Takahashi, M., and Hansen, P. J. (2012). Consequences of physiological heat shock beginning at the zygote stage on embryonic development and expression of stress response genes in cattle. *J. Dairy Sci.* 95 (6), 3080–3091. doi: 10.3168/jds.2011-4986

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146 (3), 353–358. doi: 10.1016/j.cell.2011.07.014

Shi, H., Zhu, J., Luo, J., Cao, W., Shi, H., Yao, D., et al. (2015). Genes regulating lipid and protein metabolism are highly expressed in mammary gland of lactating dairy goats. *Funct. Integr. Genomics* 15 (3), 309–321. doi: 10.1007/s10142-014-0420-1

Shwartz, G., Rhoads, M. L., VanBaale, M. J., Rhoads, R. P., and Baumgard, L. H. (2009). Effects of a supplemental yeast culture on heat-stressed lactating Holstein cows. *J. Dairy Sci.* 92 (3), 935–942. doi: 10.3168/jds.2008-1496

Silva, S. V., and Malcata, F. X. (2005). Caseins as source of bioactive peptides. *Int. Dairy J.* 15 (1), 1–15. doi: 10.1016/j.idairyj.2004.04.009

Silanikove, N., Shapiro, F., and Shinder, D. (2009). Acute heat stress brings down milk secretion in dairy cows by up-regulating the activity of the milk-borne negative feedback regulatory system. *BMC Physiol.* 9 (1), 13. doi: 10.1186/1472-6793-9-13

Spencer, J. D., Boyd, R. D., Cabrera, R., and Allee, G. L. (2003). Early weaning to reduce tissue mobilization in lactating sows and milk supplementation to enhance pig weaning weight during extreme heat stress. *J. Anim. Sci.* 81 (8), 2041–2052. doi: 10.2527/2003.8182041x

Sun, Y., Yang, Z., Zheng, B., Zhang, X. H., Zhang, M. L., Zhao, X. S., et al. (2017). A novel regulatory mechanism of smooth muscle α-actin expression by NRG-1/circACTA2/miR-548f-5p axis. *Circ. Res.* 121 (6), 628–635. doi: 10.1161/CIRCRESAHA.117.311441

Tao, S., Bubolz, J. W., Do Amaral, B. C., Thompson, I. M., Hayen, M. J., Johnson, S. E., et al. (2011). Effect of heat stress during the dry period on mammary gland development. *J. Dairy Sci.* 94 (12), 5976–5986. doi: 10.3168/jds.2011-4329

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505 (7483), 344. doi: 10.1038/nature12986

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7 (3), 562. doi: 10.1038/nprot.2012.016

Tummaruk, P., Tantasuparuk, W., Techakumphu, M., and Kunavongkrit, A. (2004). Effect of season and outdoor climate on litter size at birth in purebred Landrace and Yorkshire sows in Thailand. *J. Veterinary Med. Sci.* 66 (5), 477–482. doi: 10.1292/jvms.66.477

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., et al. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 38 (18), e178–e178. doi: 10.1093/nar/gkq622

Wang, L. I., Liu, F., Luo, Y., Zhu, L., and Li, G. (2015). Effect of acute heat stress on adrenocorticotropic hormone, cortisol, interleukin−2, interleukin−12 and apoptosis gene expression in rats. *Biomed. Rep.* 3 (3), 425–429. doi: 10.3892/br.2015.445

Wang, Y., Li, D., Wang, Y., Li, M., Fang, X., Chen, H., et al. (2019). The landscape of circular RNAs and mRNAs in bovine milk exosomes. *J. Food Composition Anal.* 76, 33–38. doi: 10.1016/j.jfca.2018.12.004

Wheelock, J. B., Rhoads, R. P., VanBaale, M. J., Sanders, S. R., and Baumgard, L. H. (2010). Effects of heat stress on energetic metabolism in lactating Holstein cows. *J. Dairy Sci.* 93 (2), 644–655. doi: 10.3168/jds.2009-2295

Wickramasinghe, S., Rincon, G., Islas-Trejo, A., and Medrano, J. F. (2012). Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* 13 (1), 45. doi: 10.1186/1471-2164-13-45

Wiersma, F., and Stott, F. M. (1974). "Response of dairy cattle to an evaporative cooled environment", in *Proceedings of the livestock environment*. (St. Joseph: American Society of Agricultural Engineers), 88–95.

Williams, A. M., Safranski, T. J., Spiers, D. E., Eichen, P. A., Coate, E. A., and Lucy, M. C. (2013). Effects of a controlled heat stress during late gestation, lactation, and after weaning on thermoregulation, metabolism, and reproduction of primiparous sows. *J. Anim. Sci.* 91 (6), 2700–2714. doi: 10.2527/jas.2012-6055

Yang, W., Du, W. W., Li, X., Yee, A. J., and Yang, B. B. (2016). Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. *Oncogene* 35 (30), 3919. doi: 10.1038/onc.2015.460

Yatoo, M. I., Dimri, M., and Sharma, M. C. (2014). Seasonal changes in certain blood antioxidants in cattle and buffaloes. *Indian J. Anim. Sci.* 84 (2), 173–176.

Yu, J., Yin, P., Liu, F., Cheng, G., Guo, K., Lu, A., et al. (2010). Effect of heat stress on the porcine small intestine: a morphological and gene expression study. *Comp. Biochem. Physiol. Part A: Mol. Integr. Physiol.* 156 (1), 119–128. doi: 10.1016/j.cbpa.2010.01.008

Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X., Chen, L. L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159 (1), 134–147. doi: 10.1016/j.cell.2014.09.001

Zhang, C., Wu, H., Wang, Y., Zhao, Y., Fang, X., Chen, C., et al. (2015). Expression patterns of circular RNAs from primary kinase transcripts in the mammary glands of lactating rats. *J. Breast Cancer* 18 (3), 235–241. doi: 10.4048/jbc.2015.18.3.235

Zhang, C., Wu, H., Wang, Y., Zhu, S., Liu, J., Fang, X., et al. (2016). Circular RNA of cattle casein genes are highly expressed in bovine mammary gland. *J. Dairy Sci.* 99 (6), 4750–4760. doi: 10.3168/jds.2015-10381

Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., et al. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287. doi: 10.1101/gr.202895.115

Zhu, S. Q., Wang, Y., Zhao, Y. L., Chen, J., Chen, C., Jiang, Y. M., et al. (2016). Identification and partial characterization of a novel circular transcript of the Tc2n gene from rat mammary gland. *Arch. Biol. Sci.* 68 (2), 257–262. doi: 10.2298/ABS151015013Z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sun, Zhang, Hu, Xie, Wang, Zhang, Chen, Luo, Wang, Jiang, Xi, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership