frontiers | Research Topics

Omics applied to livestock genetics

Edited by
Lucas Lima Verardo, Luiz Brito, Nuno Carolino and Ana Fabrícia Braga Magalhães

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-83251-807-6
DOI 10.3389/978-2-83251-807-6

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: frontiersin.org/about/contact

Omics applied to livestock genetics

Topic editors

Lucas Lima Verardo — Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

Luiz Brito — Purdue University, United States

Nuno Carolino — Instituto Nacional Investigaciao Agraria e Veterinaria (INIAV), Portugal

Ana Fabrícia Braga Magalhães — Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

Citation

Verardo, L. L., Brito, L., Carolino, N., Magalhães, A. F. B., eds. (2023). *Omics applied to livestock genetics*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-807-6

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 16 February 2023
DOI 10.3389/fgene.2023.1155611

OPEN ACCESS

EDITED AND REVIEWED BY
Martino Cassandro,
University of Padua, Italy

*CORRESPONDENCE
Lucas Lima Verardo,
lucas.verardo@ufvjm.edu.br

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

RECEIVED 31 January 2023
ACCEPTED 06 February 2023
PUBLISHED 16 February 2023

CITATION
Verardo LL, Brito LF, Carolino N and Magalhães AFB (2023), Editorial: Omics applied to livestock genetics.
Front. Genet. 14:1155611.
doi: 10.3389/fgene.2023.1155611

Editorial: Omics applied to livestock genetics

Lucas Lima Verardo[1]*, Luiz F. Brito[2], Nuno Carolino[3] and Ana Fabrícia Braga Magalhães[1]

[1]Laboratory of Animal Breeding, Department of Animal Science, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, MG, Brazil, [2]Department of Animal Sciences, Purdue University, West Lafayette, IN, United States, [3]Instituto Nacional Investigação Agraria e Veterinaria (INIAV), Oeiras, Portugal

KEYWORDS

data integration, epigenomics, farm animals, genomics, multiomics, proteomics, transcriptomics

Editorial on the Research Topic
Omics applied to livestock genetics

Since the first draft of a mammalian genome, a multitude of studies including genomics, transcriptomics, proteomics, epigenomics, and metabolomics datasets aiming to unravel the biological mechanisms influencing phenotypic expression of complex traits have been published (e.g., Portela and Esteller, 2010; Legrain et al., 2011; Costa et al., 2013; Marcobal et al., 2013; Fonseca et al., 2018; Gallagher and Chen-Plotkin, 2018). These "omics" studies have revolutionized the translation of genome to phenome research in the last two decades, including the development of important tools for the livestock sector. There are several projects, initiatives, and databases providing knowledge of genetic variations for the economically, environmentally, and socially important traits in the main livestock species. For instance, the AnimalQTLdb project (Hu et al., 2007) has curated genomic information of a large number of quantitative trait loci (QTL) identified in cattle, pigs, chicken, sheep, and other populations.

In many circumstances, the large-scale datasets generated by livestock "omics" projects have been made publicly available to researchers aiming to generate knowledge and translation tools for improving animal production and sustainability. For instance, the Functional Annotation of Animal Genomes (FAANG) project has generated datasets to decipher the function of genome segments, and it has analyzed samples from approximately 15 species, including pigs, cattle, sheep, and salmon (Giuffra et al., 2019). Moreover, the "omics" approaches can be holistically applied to improve animal breeding strategies based on biology-driven genomic predictions, besides a better understanding of the genomic background of phenotypic variability in livestock systems (Chakraborty et al., 2022).

The Research Topic titled "Omics Applied to Livestock Genetics" presents a collection of the latest findings in the area of livestock genetics based on omics approaches. Studies focusing on food-source animals such as pigs, cattle, ducks, geese, and sheep involving omics data revealed genetic information related to various relevant traits. The two most used approaches were genomics and transcriptomics in cattle and pigs (Figure 1). The results presented provide significant advancements toward understanding farm animal genetics.

In cattle studies, Silva-Vignato et al. combined SNP data from RNA-Seq and a high-density SNP panel to generate a new dataset for performing a genome-wide association analysis (GWAS) for intramuscular fat and backfat thickness in Nellore cattle. Their study

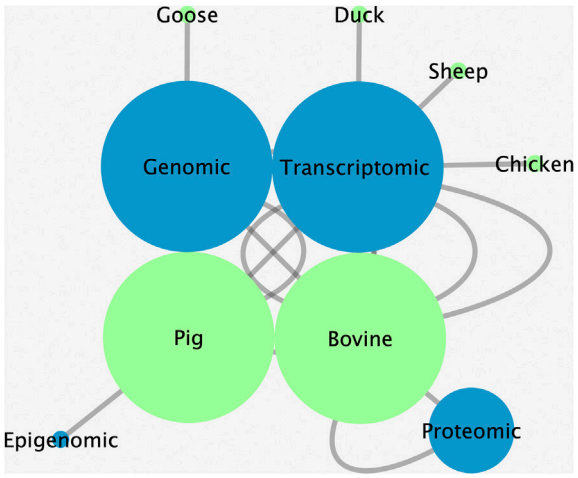


FIGURE 1
Enrichment analysis highlighting the multi-omics approaches (blue nodes) used and the most livestock species (green nodes) studied in the Research Topic. The size of the nodes corresponds to the Cytoscape network enrichment analysis, as the bigger the nodes, the more used and/or studied were the corresponding nodes.

revealed genomic regions and regulatory SNPs associated with fat deposition, including transcription factors involved in lipid metabolism-related pathways. Also integrating genomic and transcriptomic data, Liang et al. identified candidate genes for a carcass trait, the weight of *longissimus dorsi* muscle, in Huaxi cattle. After functional analysis of candidate genes and referring to other studies, key genes were suggested to be associated with body development and growth of muscle cells.

A proteomic study integrated with transcriptomic data was presented by Nguyen et al. The authors aimed to elucidate the critical proteins underlying puberty and uncover potential molecular mechanisms from the hypothalamus and pituitary gland of pre-pubertal and post-pubertal Brahman cattle. Their study identified a small number of matched transcripts and protein changes at puberty in each tissue, suggesting the need for multiple omics analyses for a better interpretation of complex biological systems. Moreover, Novais et al. applied factor analysis (FA) and Bayesian network (BN) modeling to integrate proteomic-transcriptomic data and complex traits by latent variables (production, carcass, and meat quality traits) in Nellore cattle. Their framework based on FA and BN generated new hypotheses for molecular research, by integrating different types of data and exploring hidden relationships.

Zhang et al. performed a whole-genome copy number variation detection on Suhuai (SH), Chinese Min Zhu (MZ), and Large White (LW) pigs based on next-generation sequencing data. Copy number variation regions (CNVRs) were annotated and analyzed, with some CNVRs verified by real-time polymerase chain reaction. The authors observed that SH and LW pigs are more closely related and reported annotated genes in CNVRs of each breed. Those genes were related to unique traits in each breed and thus provided important information for the identification of candidate genes for swine breeding. Shi et al. integrated lncRNA-mediated ceRNA network involved in immune regulation in the spleen of Meishan piglets. Their study collected spleen tissues from Meishan piglets at three different ages as a model, and mRNA and lncRNA transcripts were profiled. The interactions between mRNAs and lncRNAs were identified based on weighted gene co-expression network analysis, demonstrating that lncRNAs are a crucial regulatory component in mRNA. The expression of genes related to the immune response of pigs was reported, contributing to a further understanding of the mRNA and lncRNA expression in the spleen of piglets.

Moreover, aiming to explore genomic imprinting, Ahn et al. delineated spatially regulated imprinting of IGF2 transcripts, age-dependent hepatic mono-to biallelic conversion, and reorganization of topologically associating domains at the porcine H19/IGF2 locus for a better translation to human and other animal research. Using a polymorphism-based approach and omics datasets from chromatin immunoprecipitation sequencing (ChIP–seq), whole-genome sequencing, RNA-seq, and Hi-C, regulation of IGF2 during development was analyzed. Their integrative omics analyses of genome, epigenome, and transcriptome provided a comprehensive view of imprinting status at the H19/IGF2 gene cluster.

Gu et al., Gu et al. presented a transcriptome-wide study of embryonic breast muscle development in ducks and chickens, respectively. The authors performed m6A sequencing and miRNA sequencing in the breast muscle of embryos in both species. Several differentially methylated genes and differentially expressed genes were identified. They presented the first characterization of the m6A patterns in the duck transcriptome. Besides, they found that miRNAs, in conjunction with m6A modification, played a key role in the embryonic breast muscle development of Wenchang chickens.

He et al. characterized the microRNA (miRNA) and circular RNA (circRNA) expression profiles in the tail fat of sheep at 6, 18, and 30 months of age. Differentially expressed miRNAs and circRNAs were observed. Functional analysis revealed that miRNA target genes were mainly involved in cellular interactions, while the host genes of circRNAs were associated with lipid and fatty acid metabolism. miRNAs were negatively correlated with circRNAs during sheep tail fat development. Multiple ceRNA regulatory networks dominated by upregulated differentially expressed miRNAs may play a key role in this developmental process. Furthermore, Ni et al. reported whole genome sequencing analyses of a wild swan goose population. They provided a valuable data set for studies on goose genomics. These data may be useful to explore the genetic relationships between the wild swan goose and domestic goose.

In general, we observed that the main production species have been studied through omics approaches. However, multi-omics analyses are still in their infancy and the generation and sharing of multi-omics datasets will be paramount for further advancing research in this field. Functional genomic analyses and high-throughput phenotyping are crucial for providing a clearer picture of the genome-to-phenome paradigm in livestock systems. Moreover, the integration of omics technologies with phenomics into the breeding programs, which was absent from this Research Topic, may help to increase the rates of genetic progress in sustainable breeding programs.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This article was partly supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais–FAPEMIG (process APQ-01834-18).

Acknowledgments

The editors of this Research Topic thank the valuable contributions of all the authors to this Research Topic, the constructive comments and suggestions from the reviewers, and the editorial support from the Frontiers team throughout the publication process.

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.

Publisher's note

References

Chakraborty, D., Sharma, N., Kour, S., Sodhi, S. S., Gupta, M. K., Lee, S. J., et al. (2022). Applications of omics technology for livestock selection and improvement. *Front. Genet.* 13, 774113. doi:10.3389/fgene.2022.774113

Costa, V., Aprile, M., Esposito, R., and Ciccodicola, A. (2013). RNA-seq and human complex diseases: Recent accomplishments and future perspectives. *Eur. J. Hum. Genet.* 21 (2), 134–142. doi:10.1038/ejhg.2012.129

Fonseca, P. A. D. S., Id-Lahoucine, S., Reverter, A., Medrano, J. F., Fortes, M. S., Casellas, J., et al. (2018). Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. *PLoS One* 13 (10), e0205295. doi:10.1371/journal.pone.0205295

Gallagher, M. D., and Chen-Plotkin, A. S. (2018). The post-GWAS era: From association to function. *Am. J. Hum. Genet.* 102 (5), 717–730. doi:10.1016/j.ajhg.2018.04.002

Giuffra, E., Tuggle, C. K., and Faang, C. (2019). Functional annotation of animal genomes (FAANG): Current achievements and roadmap. *Annu. Rev. animal Biosci.* 7, 65–88. doi:10.1146/annurev-animal-020518-114913

Hu, Z. L., Fritz, E. R., and Reecy, J. M. (2007). AnimalQTLdb: A livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Res.* 35, D604–D609. doi:10.1093/nar/gkl946

Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., et al. (2011). The human proteome project: Current state and future direction. *Mol. Cell. proteomics* 10 (7), M111.009993. doi:10.1074/mcp.M111.009993

Marcobal, A., Kashyap, P. C., Nelson, T. A., Aronov, P. A., Donia, M. S., Spormann, A., et al. (2013). A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J.* 7 (10), 1933–1943. doi:10.1038/ismej.2013.89

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* 28 (10), 1057–1068. doi:10.1038/nbt.1685

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fgene.2022.935433

Proteomic Analysis of Hypothalamus and Pituitary Gland in Pre and Postpubertal Brahman Heifers

Loan To Nguyen[1]*, *Li Yieng Lau*[2] and *Marina Rufino Salinas Fortes*[3]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia, [2]Agency of Science, Technology and Research, Singapore, Singapore, [3]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

The hypothalamus and the pituitary gland are directly involved in the complex systemic changes that drive the onset of puberty in cattle. Here, we applied integrated bioinformatics to elucidate the critical proteins underlying puberty and uncover potential molecular mechanisms from the hypothalamus and pituitary gland of prepubertal (n = 6) and postpubertal (n = 6) cattle. Proteomic analysis in the hypothalamus and pituitary gland revealed 275 and 186 differentially abundant (DA) proteins, respectively (adjusted p-value < 0.01). The proteome profiles found herein were integrated with previously acquired transcriptome profiles. These transcriptomic studies used the same tissues harvested from the same heifers at pre- and post-puberty. This comparison detected a small number of matched transcripts and protein changes at puberty in each tissue, suggesting the need for multiple omics analyses for interpreting complex biological systems. In the hypothalamus, upregulated DA proteins at post-puberty were enriched in pathways related to puberty, including *GnRH, calcium and oxytocin signalling pathways*, whereas downregulated proteins were observed in the *estrogen signalling pathway, axon guidance and GABAergic synapse*. Additionally, this study revealed that ribosomal pathway proteins in the pituitary were involved in the pubertal development of mammals. The reported molecules and derived protein-protein networks are a starting point for future experimental approaches that might dissect with more detail the role of each molecule to provide new insights into the mechanisms of puberty onset in cattle.

Keywords: puberty, hypothalamus, pituitary gland, brahman heifers, proteomics

OPEN ACCESS

Edited by:
Ana Fabrícia Braga Magalhães, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

Reviewed by:
Kate Keogh, Teagasc Grange Animal and Bioscience Research Department (ABRD), Ireland
Guillermo Giovambattista, CONICET Institute of Veterinary Genetics (IGEVET), Argentina

***Correspondence:**
Loan To Nguyen
t.nguyen3@uq.edu.au

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

Received: *03 May 2022*
Accepted: *25 May 2022*
Published: *14 June 2022*

Citation:
Nguyen LT, Lau LY and Fortes MRS (2022) Proteomic Analysis of Hypothalamus and Pituitary Gland in Pre and Postpubertal Brahman Heifers. Front. Genet. 13:935433. doi: 10.3389/fgene.2022.935433

INTRODUCTION

Early puberty is essential for the lifetime reproductive performance of cattle (Lesmeister et al., 1973; Johnston et al., 2013). However, Brahman cattle, a breed of the *Bos indicus* sub-species, which can withstand hostile conditions in northern production systems, are often older and heavier at puberty than *Bos taurus* breeds (Chenoweth 1994). As reaching the age of puberty is an important event contributing significantly to lifetime productivity, reducing the age at puberty is a major aim for *Bos indicus* breeders for efficient herd productivity.

Like other vertebrates and humans, puberty in cattle is initiated when the hypothalamus-pituitary-ovary (HPO) axis loses its sensitivity to negative feedback effects of steroid hormones, allowing an increase in gonadotropin-releasing hormone (GnRH) secretion from GnRH neurons in

the hypothalamus. The increase then stimulates gonadotropins' secretion: luteinizing hormone (LH), and follicle-stimulating hormone (FSH) are produced and released. FSH and LH then regulate gonadal development (Day et al., 1984; Day et al., 1987; Schillo et al., 1992; Day and Anderson, 1998; Gasser et al., 2006). However, the process is controlled by multiple factors, and the complex interactions between environmental and genetic factors regulating the process are only now coming to light.

The complex genetic architecture of puberty–that is multiple variants of small effect–is probably underpinned by variants with transcriptional and post-transcriptional regulatory effects. Variations in the expression of genes or proteins in the hypothalamus or the pituitary gland, either alone or simultaneously, will affect the pubertal process. To date, several transcriptomic studies have focused on investigating candidate molecules involved in critical molecular mechanisms that drive puberty in *Bos indicus* in the HPO axis. Fortes et al., 2016 identified five transcription factors, *E2F8, NFAT5, SIX5, ZBTB38,* and *ZNF605,* with potential regulatory roles at puberty in the hypothalamus in Brahman heifers. Nguyen et al., 2017 confirmed the role of zinc finger genes in a co-expression network using ovarian data of the same Brahman heifers. Although these transcriptomic studies confirmed the complexity of puberty, the correlation between expression levels of mRNA and protein is renowned poor (Gygi et al., 1999; de Sousa Abreu et al., 2009; Carvalhais et al., 2015; Payne, 2015; Edfors et al., 2016). As such, proteomic studies are complementary information, required to advance knowledge beyond differential gene expression or co-expression analyses.

Most studies for understanding puberty-related proteins involved in Brahman heifers were performed in peripheral tissues such as the liver and adipose tissues (Nguyen et al., 2018b; 2018c). Also, Tahir and others (Tahir et al., 2019) have published on the ovarian protein abundances in pre versus post-pubertal Brahman heifers. Although a neuropeptidome was performed on the hypothalamus and pituitary gland of Brangus heifers before and after puberty (DeAtley et al., 2018), there is no previous proteome study in the hypothalamus and pituitary gland of pubertal Brahman heifers. Therefore, this present study was aimed at measuring the abundance of proteins in the hypothalamus and pituitary gland in pubertal Brahman heifers using liquid chromatography-electrospray ionisation tandem mass spectrometry (LC-ESI-MS/MS). Discovering protein abundance profiles at the different pubertal stages in the hypothalamus and pituitary gland could be useful for revealing key targets controlling the complex mechanism of puberty and complementing the previous transcriptomic studies.

MATERIALS AND METHODS

Ethics Statement

Animal use in this study was approved under animal ethics number QAAFI/279/12. Animals were housed at the Gatton Campus facilities of the University of Queensland.

Animal and Tissue Collection

The heifers used in this study and the assessment of puberty have been described by (Fortes et al., 2016; Nguyen et al., 2017; Nguyen et al., 2018a; Lau et al., 2020). In brief, these heifers were of the same age with no difference in body weight or body condition scores. Pubertal status was defined by the presence of the first *corpus luteum* (CL) and progesterone concentration. Postpubertal heifers were euthanized by stunning with a nonpenetratng captive bolt through the parietal bone of the head around day 15 after observing the first CL. The nonpenetrating captive bolt methodology was used for protecting the integrity of the lower brain tissues as previously described (Cánovas et al., 2014). A pre-pubertal heifer was then randomly selected to pair with a postpubertal heifer on slaughter day. Plasma progesterone concentrations were 0.4 ± 0.2 ng/ml and 2.0 ± 0.7 ng/ml in pre and post-pubertal heifers, respectively.

After euthanasia, the hypothalamus tissue (spanning from the preoptic to the arcuate nucleus) and pituitary gland (including anterior and posterior pituitary gland) tissues were quickly harvested, frozen in liquid nitrogen, and stored at −80°C until tissue processing.

Protein Sample Preparation

Protein extraction and digestion of heifers' hypothalamus and pituitary tissues were performed as previously described (Tan et al., 2014; Nguyen et al., 2018b; Nguyen et al., 2018c). Briefly, the hypothalamus and pituitary tissues were manually homogenized into smaller fragments using rigid aluminium foil to protect the tissue and a hammer to "pulverize" the sample. This procedure was performed over a dry-ice bed to preserve the sample and randomize the fragments, ensuring that each sample was representative of the entire tissue. The resulting tissue fragments were transferred into Eppendorf® lobind microcentrifuge tubes (Sigma-Aldrich) where lysis solution (7 M urea, 2 M thiourea, 4%SDS, 10 mM DTT and 1 mM PMSF) was added. The fragments in the solution were sonicated at power level 4 for 10 s. Subsequently, the homogenate was vortexed vigorously for 1 h at 30°C. Following that, 25 mM acrylamide was added to the samples and subsequently incubated for 1 h at 30°C. Next, 5 mM DTT was added to samples in order to quench excess acrylamide. Four volumes of methanol: acetone (1:1) were added, and the samples precipitated overnight at −20°C. The precipitates were subsequently dissolved by adding 50 mM ammonium acetate, and the protein concentrations were measured with a Nanodrop (Thermo Scientific). After measuring concentration, about 100 μg of protein were transferred into a 10 kDa Amicon Ultra 0.5 centrifugal filter device (Merck Millipore). This filter device was then inserted into a collection tube and centrifuged at maximum speed for at least 30 min. Protein was diluted in 50 mM ammonium bicarbonate and then again centrifuged at maximum speed for at least 30 min. Trypsin was used as a protease for digestion of protein solution followed by incubation overnight at 37°C. Peptides were desalted by C-18 Zip-tip (adapted from Millipore procedure) and stored at −20°C until analysis.

Mass Spectrometry and Data Analysis

The LC-ESI-MS/MS was performed using a Prominence nanoLC system (Shimadzu) and TripleToF 5600 mass spectrometer with a Nanospray III interface (SCIEX) as previously described (Xu et al., 2015). Peptides were separated using a 70-min LC gradient. MS-TOF scan was performed. We performed information-dependent acquisition (IDA) of top peptides for one randomly chosen pre-pubertal sample and one randomly chosen post-pubertal sample. Subsequently, sequential window acquisition of all theoretical mass spectra (SWATH) was performed on all samples using the prepared IDA library for each group.

Protein Pilot v5.0.1 (ABSCIEX) was then utilized for peptide identification. The bovine protein database was used for peptide mapping and retrieved from Uniprot (www.uniprot.org; 43,813 entries assigned to *Bos taurus*). For subsequent analyses, identified peptides with more than 99% confidence and a false discovery rate (FDR) of less than 1% were used. Ion libraries were used for SWATH analyses. The abundance of proteins and peptides was computed by PeakView v2.1 software (ABSCIEX). The differential expression analysis between the two grops of pre and post-pubertal heifers as performed using MSstats (v2.6) in R (Choi et al., 2014). For controlling the false discovery rate, the *p*-values were adjusted using the Benjamini and Hochberg's approach. Proteins with an adjusted *p*-value < 0.01 were assigned as differentially abundant.

Transcriptomic Data

The hypothalamus and pituitary gland transcriptomic data were obtained from Fortes et al., 2016 and Nguyen et al., 2017. These transcriptomic studies used the same tissues harvested from the same heifers at pre- and post-puberty. The mRNA and protein pairs were identified from sets of expressed genes (Fortes et al., 2016; Nguyen et al., 2017) and set of abundant proteins (current study). The Pearson's correlation coefficient (*r*) was used to compute the correlations between expression levels of mRNAs and the abundance of proteins in the hypothalamus and pituitary gland from pre- and postpubertal Brahman heifers. The cor() function in R was used in this calculation. The unmatched mRNA-proteins were not included in this analysis.

Functional Enrichment Analysis of DA Proteins

The gene ontology analysis of differentially abundant (DA) proteins was obtained through STRINGv10 system (http://string-db.org). The DA proteins were classified according to biological process (BP), cellular component (CC), molecular function (MF) and pathways. Further, Cytoscape software (http://cytoscape.org/) was used to retrieve biological terms for the protein interaction networks. In order to determine significant terms and pathways, the terms having a corrected *p*-value (FDR) < 0.05 were considered.

RESULTS

Protein Identification, Quantification, and Differential Abundance

In total, 765 proteins were identified and quantified in the hypothalamus. Of this total, 275 differentially abundant (DA) proteins were identified (adjusted *p*-value < 0.01) in the hypothalamus libraries, of which 120 were significantly downregulated and 155 were significantly upregulated at post-puberty (adjusted *p*-value < 0.01) (**Figures 1A**, **2A** and **Supplementary Table S1**).

In the pituitary libraries, a total of 715 proteins were identified and quantified. Of these, 186 were DA proteins, where 96 proteins were significantly decreased, and 90 proteins were significantly increased at post-puberty (adjusted *p*-value < 0.01) (**Figures 1B**, **2B** and **Supplementary Table S1**).

The extent of DA proteins between pre and post-puberty in each tissue was visualized in volcano plots (**Figures 2A,B**). Protein IDs, corresponding gene names and fold change (log_2FC) information are listed in **Supplementary Table S1**.

Derived Correlations Between mRNA Levels and Protein Abundances

The matching pairs of expressed mRNA and proteins in each tissue were used for comparison because of the relatively small number of matched DE mRNA and DA proteins. In the hypothalamus, 641 out of 765 identified proteins corresponded to mRNA transcripts discovered in the hypothalamic transcriptome data (Fortes et al., 2016) (**Figure 1A** and **Supplementary Table S2**). The correlation between 641 protein and transcript pairs in the hypothalamus was positive but insignificant, with a correlation coefficient of 0.05 (*p*-value = 0.16). Among these pairs, only eight DE genes were DA proteins, namely *TUBB2B, OMG, SPTAN1, RAP1A, SSBP1, PLCB1, EIF3J* and *H2AFX* (**Table 1**). Six of these eight DA proteins showed the same direction of regulation as their mRNAs.

When compared the proteins in the pituitary gland with its corresponding mRNA data (Nguyen et al., 2017), 648 out of 715 mRNA-proteins pairs were identified (Fortes et al., 2016) (**Figure 1B** and **Supplementary Table S2**). The correlation between transcript and protein in pituitary for 648 proteins was minimal (r = 0.006; *p*-value = 0.8). Among these pairs, five DA proteins and DE genes were in common. They were: *IGFBP2, TAGLN, CHGB, ENO1* and *HIST2H2AC* (**Table 1**). Four of these five proteins showed the same direction of regulations as their mRNAs.

Gene Ontology Enrichment Analysis

In order to evaluate the major biological processes influencing puberty onset in Brahman cattle, GO enrichment was performed to examine the functional characteristics of DA proteins between pre and post-pubertal samples in each tissue (**Supplementary Tables S3, S4**).

As shown in **Figure 3**, DA proteins in the hypothalamus were predominantly enriched for "organic substance metabolic process" (31%), "single-organism metabolic process" (27%), "oxidation-reduction process" (14%) and "nervous system

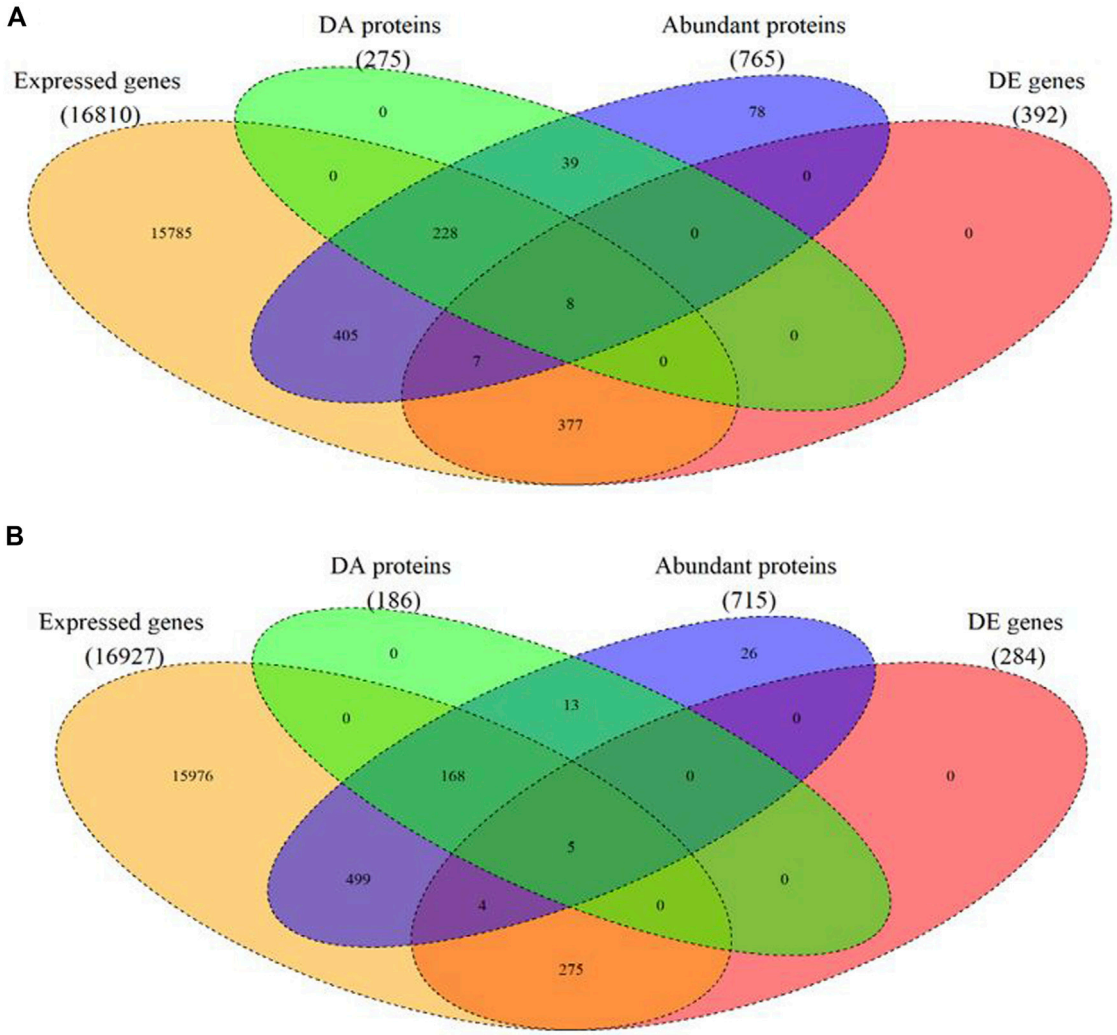


FIGURE 1 | Venn diagram of expressed mRNAs and proteins as well as differentially expressed (DE) mRNA and differentially abundant (DA) protein in each tissue. **(A)** in the hypothalamus tissue, **(B)** in the pituitary gland.

development" (12%), in the BP category of GO terms. As expected, DA proteins were involved in the peptide metabolic process and brain development (including development of neurons, glial cells and axons). These DA proteins were also annotated to the terms "response to oxidative stress," "glutamine metabolic and catabolic process," "energy derivation by oxidation of organic compounds," and "regulation of stress-activated MAPK cascade" (FDR <0.05). MF analysis revealed DA proteins were connected to 89 enriched GO terms, including proteins involved in binding (38%) and catalytic activity (34%). There were 116 enriched GO terms found in the CC category (FDR <0.05), and the majority of DA proteins were located in the cytoplasmic (51%) and intracellular (45%) space.

According to the GO annotations of DA proteins from the pituitary gland (**Figure 4**), among 214 enriched BP terms, the majority of DA proteins were classified in the single-organism cellular process (39%), cellular process (38%), metabolic process (35%) and biological regulation (30%). Enriched MF terms were 39 in the analysis of the pituitary gland (FDR <0.05). Hydrolases, oxidoreductases and endopeptidase were the main protein classes identified in this MF anotation. Moreover, the CC terms of pituitary DA proteins were assigned to the organelle, membrane-bounded organelle, cytoplasm, intracellular and extracellular regions.

KEGG Pathway Analysis

In order to investigate the biological pathways enriched in response to puberty onset in Brahman heifers, a KEGG pathway enrichment analysis was performed.

In the hypothalamus, comparison of pre- and post-pubertal samples, yielded a total of 275 DA proteins mapped to 67 enriched pathways (**Supplementary Table S5**). Among enriched pathways, known pathways that are related to puberty were also identified. The enriched pathways included the *estrogen signalling pathway* (FDR = 0.002), *axon guidance* (FDR = 0.006), *glutamatergic*

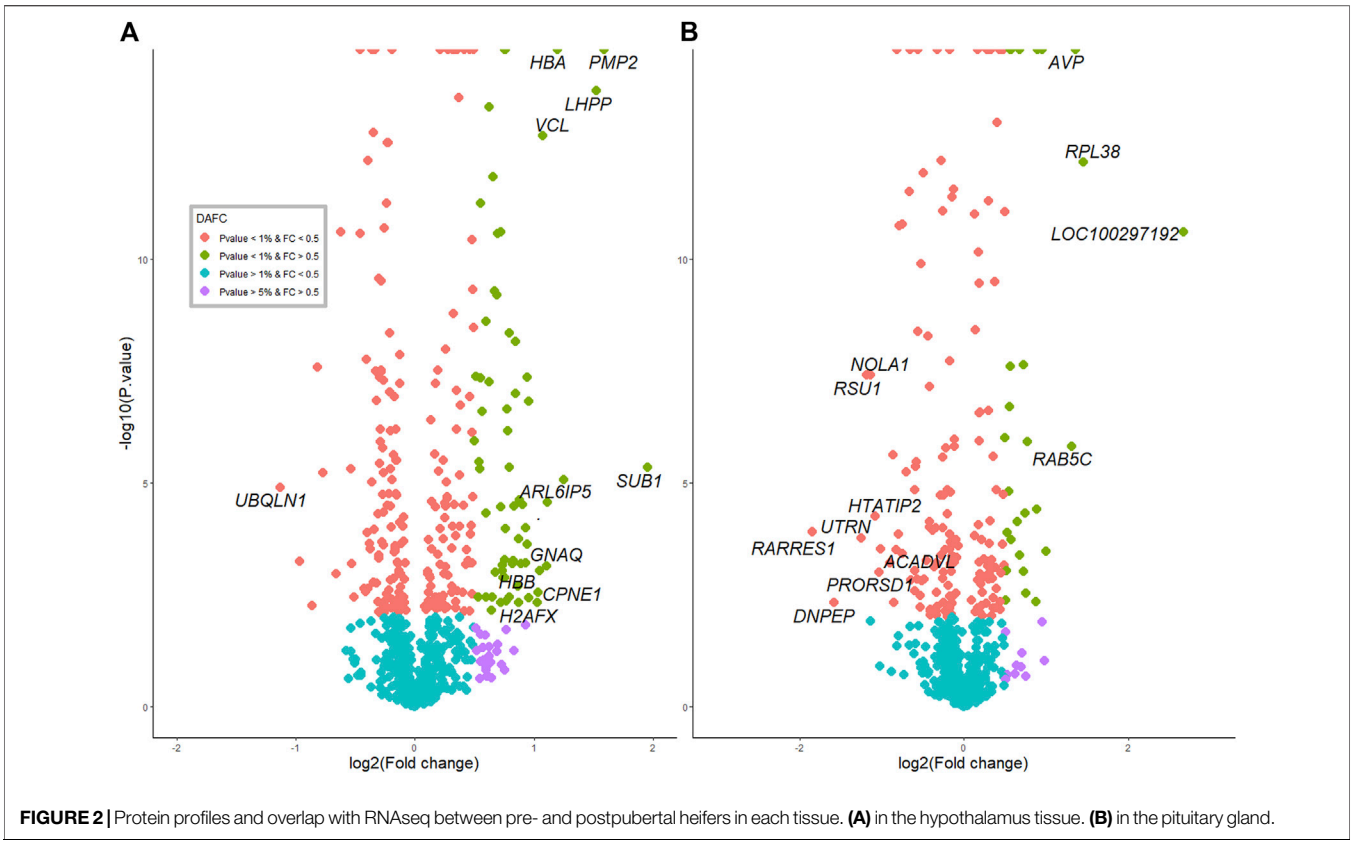


FIGURE 2 | Protein profiles and overlap with RNAseq between pre- and postpubertal heifers in each tissue. **(A)** in the hypothalamus tissue. **(B)** in the pituitary gland.

TABLE 1 | List of common differentially expressed genes and abundant proteins in the hypothalamus and pituitary studies.

Tissue	Ensembl ID	Uniprot ID	Gene symbol	FC_mRNA	FC_protein
HYP	ENSBTAG00000004093	Q6B856	*TUBB2B*	−0.203	−0.394
	ENSBTAG00000025213	Q0IIH3	*OMG*	0.199	−0.297
	ENSBTAG00000015327	E1BFB0	*SPTAN1*	−0.131	−0.117
	ENSBTAG00000014710	P62833	*RAP1A*	0.113	0.087
	ENSBTAG00000010931	F1N1S0	*SSBP1*	0.189	0.321
	ENSBTAG00000008338	P10894	*PLCB1*	−0.319	0.779
	ENSBTAG00000000359	G8JKV2	*EIF3J*	0.119	0.845
	ENSBTAG00000038047	Q17QG8	*H2AFX*	0.199	1.026
PIT	ENSBTAG00000005596	F1N2P8	*IGFBP2*	−0.615	−0.452
	ENSBTAG00000007196	V6F957	*TAGLN*	−0.405	−0.698
	ENSBTAG00000011782	P23389	*CHGB*	−0.266	−0.653
	ENSBTAG00000013411	F1MB08	*ENO1*	−0.126	−0.170
	ENSBTAG00000032456	F2Z4I6	*HIST2H2AC*	0.680	−0.338

FC, means fold change; HYP, means Hypothalamus; PIT, means Pituitary.

synapse (FDR = 0.0001), *GABAergic synapse* (FDR = 0.005), *GnRH signalling pathway* (FDR = 0.02), *oxytocin signalling pathway* (FDR = 0.001), *calcium signalling* (FDR = 0.01) and *PPAR signalling pathway* (FDR = 0.05). DA proteins involved in the *GnRH signalling pathway* were highlighted in **Figure 5**. In addition, DA proteins were also involved in other pathways such as *oxidative phosphorylation* (FDR = 6.6×10^{-12}), *the tricarboxylic acid cycle* (TCA, FDR = 3.58×10^{-9}), *non-alcoholic fatty liver disease* (NAFLD, FDR = 1.08×10^{-5}), *glycolysis and gluconeogenesis* (FDR <0.05), *thyroid hormone synthesis* (FDR = 0.05), *metabolism of amino acids* and *carbohydrate metabolism* (FDR <0.05). The DE gene and corresponding DA protein, *PLCB1*, was annotated to four enriched pathways: *GnRH signalling*, *estrogen signalling*, *calcium signalling* and *oxytocin signalling*. In short, pathways related to neuronal-hormonal signalling and metabolism were significant for the DA proteins in the hypothalamus (**Table 2**).

In the pituitary gland, a total of 186 DA proteins were assigned to 12 KEGG pathways (FDR <0.05). Among these enriched pathways, the *ribosome pathway* represented the largest number of DA proteins (FDR = 0.0002). *Complement and*

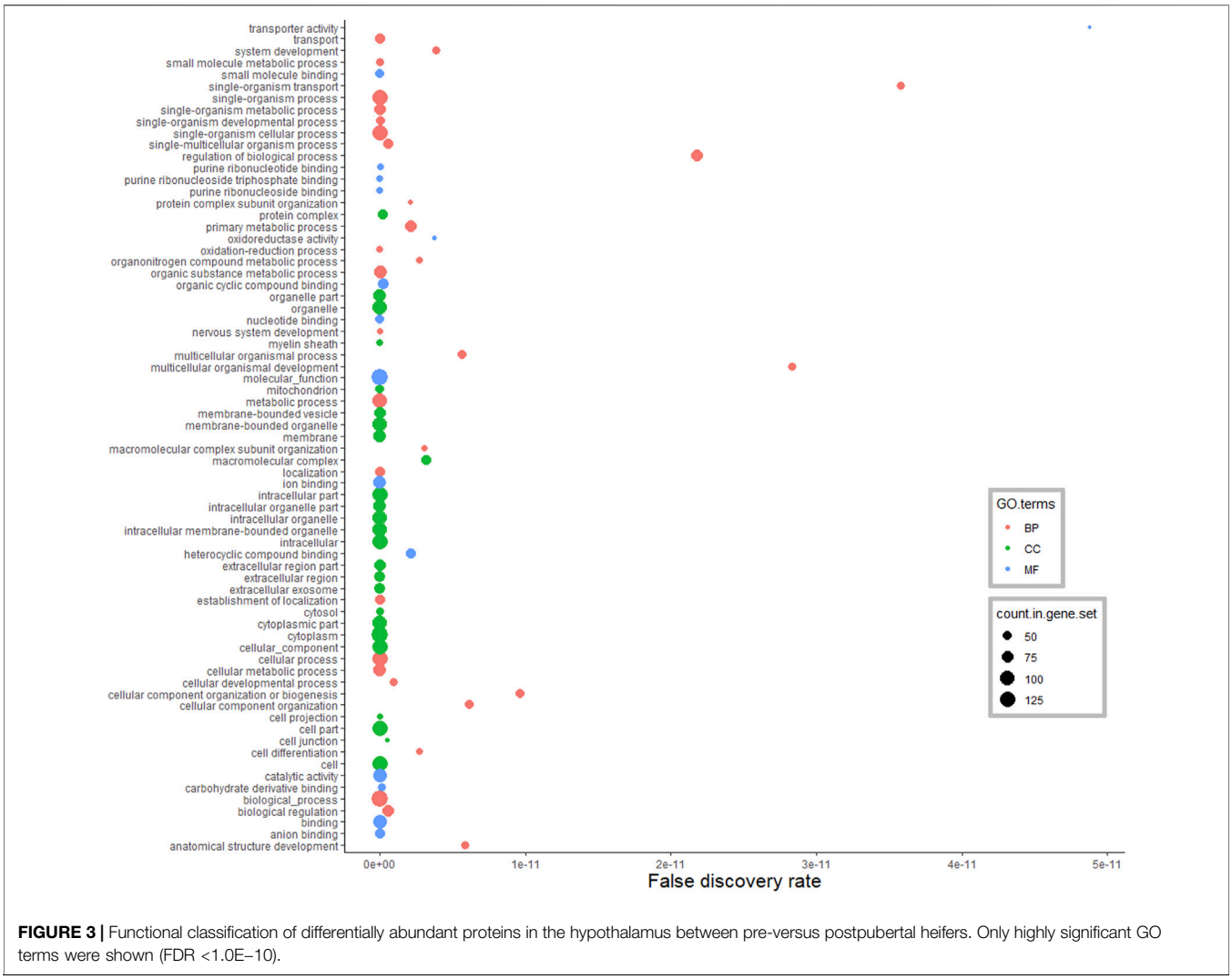


FIGURE 3 | Functional classification of differentially abundant proteins in the hypothalamus between pre-versus postpubertal heifers. Only highly significant GO terms were shown (FDR <1.0E−10).

coagulation cascades, *glycolysis and gluconeogenesis*, *biosynthesis of amino acids*, and *focal adhesion* were also overrepresented (**Figure 6**).

Protein -Protein Interaction Network Analysis

The relationship among DA proteins was further investigated in a network analysis. The hypothalamus network contained 181 nodes (proteins) with 486 edges (interactions) (**Figure 7**). Heat shock protein 90 alpha family class A member 1 (*HSP90AA1*) had the highest number of connections in the network (34 edges), followed by malate dehydrogenase 2 (*MDH2*) and albumin (*ALB*) with 23 and 21 connections, respectively. Of note, *HSP90AA1* contributes to the *estrogen signalling pathway*, according to the KEGG pathway analysis (**Table 2**). In addition, *KRAS* protein, which plays a role in *GnRH signalling*, *axon guidance*, *estrogen signalling* and *oxytocin signalling*, interacted with 13 proteins in the network (**Table 2** and **Figure 7**). PLCB1, a DE gene and DA protein, had nine connections in the network: *GNAQ, GNAI2, PIP4K2A, GNAO1, SYNJ1, GNB2, PRKCG, CAMK2D, CAMK2A* (**Table 2** and **Figure 7**). These genes that were hubs in the network analysis and were associated with enriched pathways of biological relevance for puberty might be essential drivers of puberty in cattle.

The interaction network from the pituitary DA proteins comprised 113 nodes with 254 edges (**Figure 8**). Among these nodes, ribosome-related proteins seem to form a dominant cluster in the network. The protein RPS5 had 17 connections within the network. A common DE gene and a DA protein, *ENO1*, was found in the network with 10 connections. This protein is involved in carbon metabolism, biosynthesis of amino acids and glycolysis/gluconeogenesis, as identified by the KEGG pathway analysis of DA proteins in the pituitary between pre- and post-pubertal Brahman heifers.

DISCUSSIONS

High-throughput techniques such as transcriptomic and proteomic studies are potent approaches to understanding

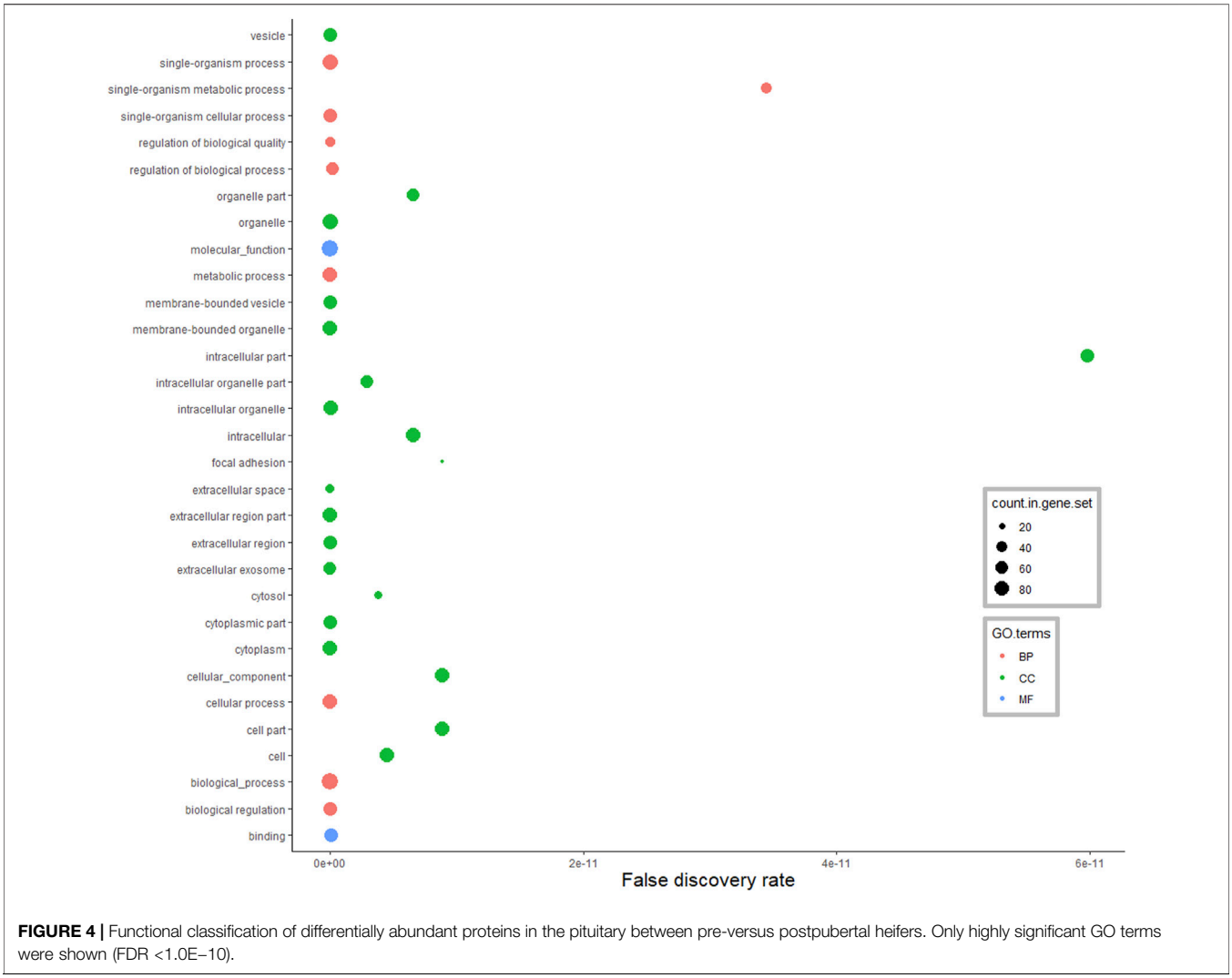


FIGURE 4 | Functional classification of differentially abundant proteins in the pituitary between pre-versus postpubertal heifers. Only highly significant GO terms were shown (FDR <1.0E−10).

gene expression and regulation. Mowever, no single approach can stand alone if we are to understand the fundamental biology underlying complex traits. Analyses at multiple levels are needed to interpret complex biological systems. In this study, the correlation between protein abundance and mRNA expression in the hypothalamus and pituitary gland was insignificant. To date, reports on the correlation between mRNA and proteins that use high-throughput approaches are scarce. Yeast, bacteria and human cancer data reports found limited correlations between mRNA and protein levels (Anderson and Seilhamer, 1997; Gygi et al., 1999; Chen et al., 2002; Griffin et al., 2002; Ghaemmaghami et al., 2003; Greenbaum et al., 2003; Washburn et al., 2003; Tian et al., 2004; Nie et al., 2006; Jayapal et al., 2008; de Sousa Abreu et al., 2009; Haider and Pal, 2013; Bai et al., 2015; Carvalhais et al., 2015; Payne, 2015; Edfors et al., 2016). A weak correlation of 0.05 was reported for 17 DE genes and corresponding protein levels analysed in the bovine mammary gland (Dai et al., 2017). The weak correlation between mRNA and protein expression levels may result from methodological constraints and biological factors, such as translational regulation and protein *in vivo* half-lives (Maier et al., 2009; Vogel and Marcotte, 2012; Haider and Pal, 2013). Additional data on factors affecting protein levels and in-depth studies, can help address this conundrum.

Hypothalamus transcriptome and proteome analyses in pre- and post-pubertal Brahman heifers revealed eight DE genes that are also DA proteins. Among these DA proteins, tubulin beta 2B class IIb (*TUBB2B*) is involved in neuronal migration and axonal guidance that are prominent brain remodeling mechanisms during pubertal development (Bond et al., 1984; Romaniello et al., 2015). A study in female rats noted that the neurotrophic effects of estrogen on the central nervous system at the onset of puberty were partly mediated by the increase of tubulin beta class II (Rogers et al., 1993). Further, the significant increase of *TUBB2B* and *TUBB2A* mRNA expression in low progesterone endometrium was reported in heifers (Forde et al., 2012). Our study revealed a decreased abundance of *TUBB2B* post-puberty, suggesting the correlation between TUBB2B and steroid hormones. High expression of *TUBB2B* in pre-pubertal Brahman heifers is proposed to play a part in

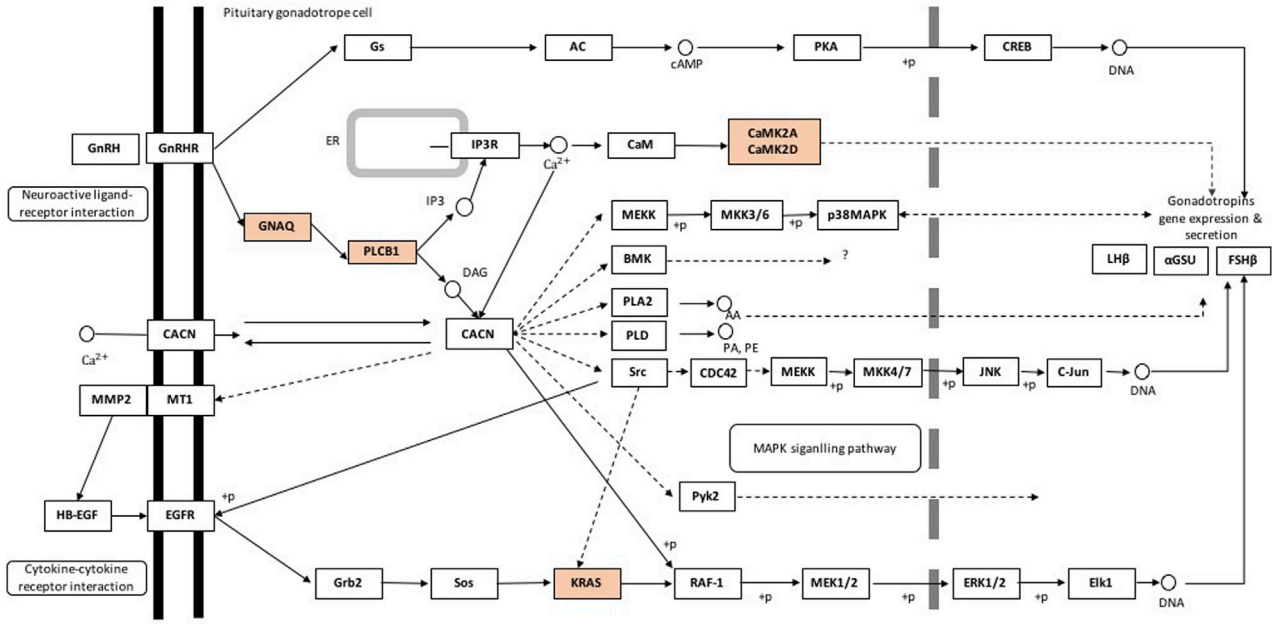


FIGURE 5 | Gonadotropin-releasing hormone signalling pathway containing differentially abundant proteins in the hypothalamus of pre- and postpubertal brahman heifers. Pathway components representing differentially abundant proteins are in orange. Pathway was adapted from the KEGG database (http://www.genome.jp/kegg-bin/show_pathway?bta04912, accessed: 11/03/2018).

TABLE 2 | Enriched pathways related to differentially abundant proteins in heifer puberty in the hypothalamus.

Pathway name	Count	False discovery rate	Proteins involved
GnRH signalling pathway	5	0.02	**GNAQ**, **PLCB1**, **CAMK2A**, **CAMK2D**, KRAS
Calcium signalling pathway	8	0.01	**SLC25A6**, **ATP2B3**, **GNAQ**, **PLCB1**, **PRKCG**, **CAMK2A**, **CAMK2D**, **ATP2B4**
Oxytocin signalling pathway	9	0.001	**EEF2**, **PRKCG**, **CAMK2A**, **GNAQ**, **CAMK2D**, **PLCB1**, GNAI2, GNAO1, KRAS
Axon guidance	7	0.006	GNAI2, RAC1, KRAS, CDK5, DPYSL2, **DPYSL5**, **CFL1**
GABAergic synapse	6	0.004	**GAD2**, **GLUL**, **PRKCG**, GNAO1, GNAI2**,** GNB2
Estrogen signalling pathway	7	0.002	HSPA8, GNAO1, KRAS, GNAI2, HSP90AA1**, PLCB1**, **GNAQ**
PPAR signalling pathway	4	0.05	DBI, ILK, **APOA2**, FABP7

Low and high abundant proteins at post-puberty (adjusted p*-value < 0.01) appear in normal and bold type, respectively.*

remodeling the brain and how it responds the feedback of ovarian hormones. Changes to the central nervous system are intrincical to pubertal development.

Another protein among the eight that were DA that were also DE genes in the hypothalamus is spectrin alpha nonerythrocytic 1 (*SPTAN1*). The *SPTAN1* gene is known to play a vital role in brain development and epileptic encephalophathy (Wang et al., 2018). This protein was identified as an upregulated gene in the mammary epithelial cells of pre-pubertal female mice compared to post-pubertal mice (Pal et al., 2017). These results in mice seem contrasting to ours, as *SPTAN1* was downregulated in post-pubertal cattle. Species or tissue differences and the definition for puberty (and therefore timing of sampling) could explain such discrepancies. Nonetheless, SPTAN1 is emerging as a candidate for future research targeting puberty machanisnsm in mammals.

Perhaps the Phospholipase C beta 1 (*PLCB1*) was the DA protein, which is also a DE gene, with more evidence of its role in puberty. *PLCB1* plays a crucial role in reproductive physiology. Genome-wide association studies in pigs reported the contribution of *PLCB1* to growth at puberty onset and in the gonadotropin signalling pathway (Meng et al., 2017). Further, *PLCB1* disruption resulted in infertile mice with pleiotropic reproductive defects (Filis et al., 2013). PLCB1 is a crucial factor modulating GPCR signalling, which controls reproductive physiology in mice (Jiang et al., 1997; Kim et al., 1997; Xie et al., 1999; Bohm et al., 2002; Ballester et al., 2004; Filis et al., 2013). Our results serve as evidence that PLCB1 may have a role in puberty onset in *Bos indicus* cattle: it interacts in the predicted network with other genes (GNAQ, GNAI2, PIP4K2A, GNAO1, SYNJ1, GNB2, PRKCG, CAMK2D, CAMK2A) related to GnRH signalling and other hormonal and neuronal signalling pathways known to be involved in the feedback mechanisms driving the activation of HPO axis. More importantly, PLCB1 was one of the upregulated DA proteins at post-puberty involving in *GnRH signalling*, *estrogen signalling*, *calcium signalling* and *oxytocin signalling*. In hypothalamic neurons, the pulsatile

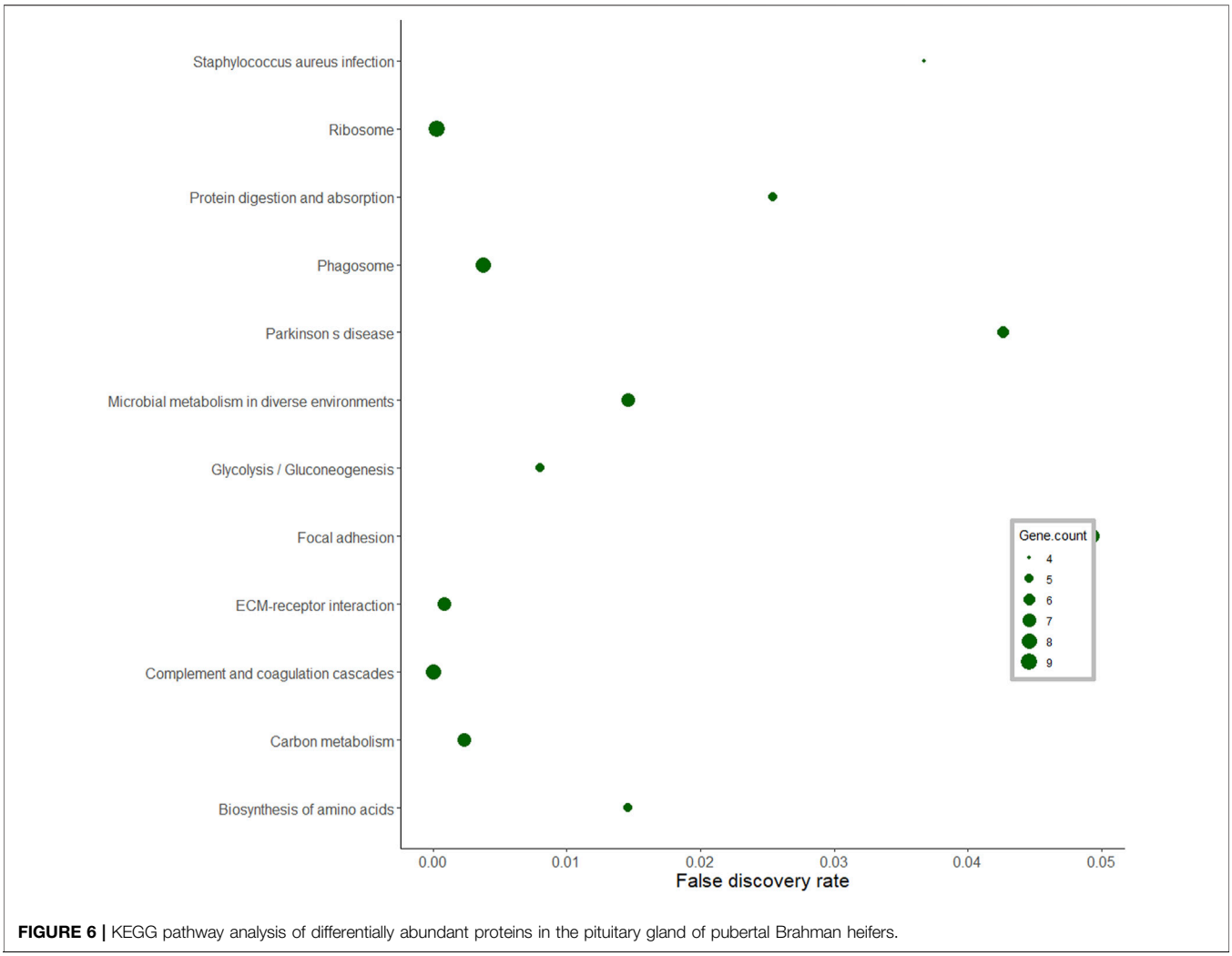


FIGURE 6 | KEGG pathway analysis of differentially abundant proteins in the pituitary gland of pubertal Brahman heifers.

release of GnRH is depended on voltage-gated calcium influx (Krsmanović et al., 1992). Therefore, it is logical to propose that the increase of GnRH release at the hypothalamus of Brahman heifers may be initiated *via* oxytocin signalling, *via* PLCB1.

Pituitary transcriptome and proteome analyses revealed five differentially expressed genes: *IGFBP2, TAGLN, CHGB, ENO1* and *HIST2H2AC,* significantly less abundant post-puberty. Among these, *IGFBP2* was the DA and DE gene with mounting evidence of its roles in puberty. *IGFBP2* is a predominant insulin growth factor binding protein synthesised and secreted in the anterior pituitary during the pre-ovulatory and early luteal phase in beef cattle (Funston et al., 1995; Roberts et al., 2001). In the anterior pituitary, the expression of *IGFBP2* fluctuated with changes in the estrous cycle that were associated with serum progesterone (Funston et al., 1995). In addition, estrogen signalling increased *IGFBP2* expression in the anterior pituitary of ewes, cattle, pigs and rats (Michels et al., 1993; Clapper et al., 1998; Roberts et al., 2001; Rempel and Clapper, 2002). Roberts et al. (2001) reported that IGFBPs levels in the anterior pituitary decreased from pre-ovulatory to early luteal development in beef cattle. Consistent with Robert et al. (2001) observations, the mRNA expression of *IGFBP2* was decreased in the transcriptome study (Nguyen et al., 2017) and the current study. In short, we observed a lower abundancy of *IGFBP2* in the bovine pituitary gland at the luteal phase post-puberty. The release of *IGFBP2* in the pituitary gland was stimulated by gonadotropin-releasing hormone, as demonstrated by Robert et al. (2001). Also, polymorphisms in *IGFBP2* were associated with age at puberty in Brahman cattle (Fortes et al., 2013b).

Chromogranin B (*CHGB*), another DA protein that was also a DE gene from the pituitary studies, belongs to a chromogranin protein family, which has been noted to affect the secretion and storage of FSH and LH at different periods of estrous cycle in sheep (Crawford and McNeilly, 2002). A study of granin-gonadotropin interactions in sheep observed a decrease of *CHGB* mRNA level after the pre-ovulatory LH surge (Crawford and McNeilly, 2002). In agreement with the sheep study, mRNA expression of *CHGB* decreased at post-puberty in the pituitary transcriptome of Brahman heifers (Nguyen at el., 2017). Additionally, the neuropeptide *CHGB* was selected as one of the fertility-related neuropeptides in the pituitary gland

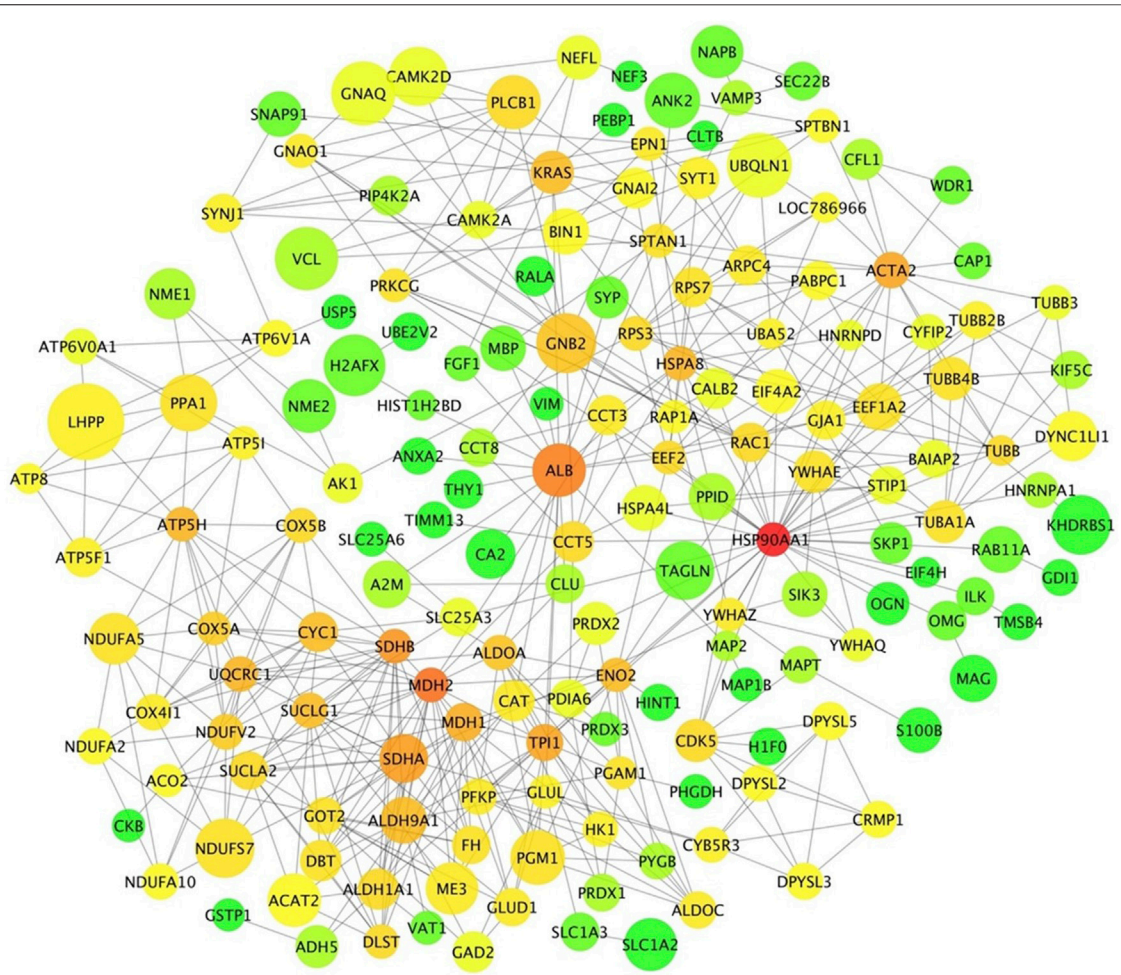


FIGURE 7 | Protein-protein interaction network in the hypothalamic tissue. Each node represents a protein. Edges illustrate the interaction between proteins. Node size represents the level of differential abundance measured for each protein in absolute terms, where bigger nodes represent higher fold changes. Node colour range from green to red represent low to high connectivity in the network for each specific protein.

neuropeptidome of pre- and post-pubertal Brangus heifers (DeAtley et al., 2018). The neuropeptidome study of Brangus heifers also reported higher levels of *CHGB* neuropeptide products in pre-pubertal pituitary tissue (DeAtley et al., 2018). This study observed a lower abundance of *CHGB* protein after puberty in Brahman heifers. Although measuring neuropeptides is different from measuring protein levels, both studies point to the importance of CHGB in the pituitary glands during pubertal development in cattle.

The DA protein α-enolase (*ENO1*) is one of the three isoforms of glycolytic enzymes responsible for catalysing the conversion of 2-phosphoglycerate to phosphoenolpyruvate (Pancholi, 2001). A study in the ovaries of pre-laying and laying geese suggested that *ENO1* may regulate reproductive function in female geese (Ji et al., 2014; Kang et al., 2014). The overexpression of *ENO1* in granular cells induced the mRNA expression of FSH and LH receptors (Ji et al., 2014). Activation of these receptors is necessary for the hormonal functioning of pituitary glycoprotein hormones (LH and FSH). Pituitary FSH and LH concentrations decreased at puberty in heifers (Desjardins and Hafs, 1968). Our current and transcriptome studies (Fortes et al., 2016; Nguyen et al., 2017) found lower mRNA and protein expression of *ENO1* post-puberty. On the contrary, single-cell RNA profiling found *ENO1* to be down-regulated in a pre-puberty gene network of mouse mammary glands (Pal et al., 2017). It is apparent that *ENO1* impacts female puberty, but its precise function needs to be considered in terms of tissue and species-specific.

The DA protein coded by the DE gene *HIST2H2AC* in the pituitary, a member of the histone 2A family, was mapped to a genetic locus reported linking puberty timing and pubertal height growth in humans (Cousminer et al., 2013). A genetic correlation between age at puberty and high height in Brahman heifers was also reported (Vargas et al., 1998). Differential expression of *HIST2H2AC* at mRNA and proteins level between pre- and post-pubertal heifers proposes it has a role in pubertal development. This candidate gene could be mined for mutations that could be tested for their effects on height and puberty in cattle.

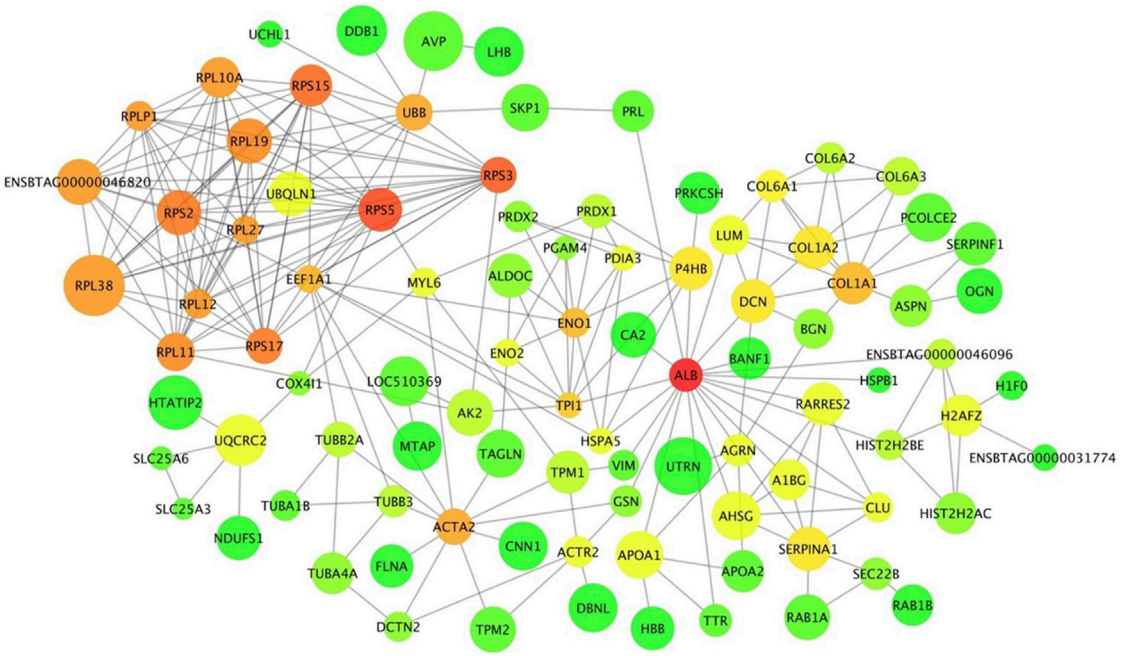


FIGURE 8 | Protein-protein interaction network of differentially abundant proteins between pre- and postpubertal heifers in the pituitary gland. Each protein is a node and protein-protein interactions are illustrated as edges that link nodes. The size of node represents the increase or decrease in protein abundance post-puberty. Node colour represents the connectivity of a protein within the network; it is ranged from green to red for low to high connectivity.

The activation of HPO axis culminates with the pulsatile release of GnRH from the hypothalamus. GnRH triggers the synthesis and secretion of LH and FSH in the pituitary gland, which stimulates the production of gonadal hormones leading to ovulation (Bliss et al., 2010). Both positive and negative feedback at several levels regulate HPO function. One positive feedback on the activation of GnRH secretion is commenced by GnRH receptor (GnRHR) - G-protein α_q subunit ($G\alpha_{q/11}$), which induces increased intracellular calcium (Khadra and Li, 2006). In agreement with this positive feedback, our proteomic study in the hypothalamus revealed the up-regulation of the calcium signalling pathway (FDR = 0.05) and GnRH signalling pathway (FDR = 0.02). G protein subunit alpha q (GNAQ), phospholipase C beta 1 (PLCB1), calcium/calmodulin-dependent protein kinase II alpha (CAMK2A), and calcium/calmodulin-dependent protein kinase II delta (CAMK2D) showed increased abundance at post-puberty in Brahman heifers and were associated with the upregulated pathways. As a result, enrichment analysis of DA proteins from the hypothalamus confirmed the involvement of these proteins on the onset of puberty *via* multiple pathways.

Among enriched GO terms in the pituitary libraries, six of these 11 GO terms, including regulation of multicellular organismal process, extracellular region part, extracellular matrix, proteinaceous extracellular matrix, calcium ion binding and extracellular matrix structural constituent were also found as GO terms involved in the onset of puberty in goat and rats (Gao et al., 2018). These findings and the referenced literature suggest the involvement of these GO terms in the regulation of puberty in goats, rats and Brahman heifers. It seems that many of the mechanisms of female puberty are conserved across mammals.

The interaction network in the hypothalamus revealed three DA proteins as highly connected hubs: HSP90AA1, MDH2 and ALB. The protein HSP90AA1 was involved in estrogen signalling, and it interacted with other proteins of the PPAR and oxytocin pathways (ILK, EEF2 and PRKCG). Recently, a transcriptomic study in pre- versus post-pubertal mammary epithelial cells noted *HSP90AA1* as a DE gene (Pal et al., 2017). Even though there is little evidence of a relationship between MDH2 and female reproduction, MDH2 was characterised as a protein marker for male fertility (Kwon et al., 2015a; Kwon et al., 2015b). A study in Egyptian boys suggested ALB status involved in each puberty stage (Cole et al., 1982). The increase of albumin excretion rate in the urinary tract was significantly associated with puberty in non-diabetic children and adolescents (Bangstad et al., 1993). Bovine serum albumin–estrogen compounds impacted GnRH1 neuronal activity (Temple and Wray, 2005). The inferred network predicted interactions between ALB and proteins involved in estrogen (HSPA8) and oxytocin (EEF2) signalling.

Gao et al. (2018) performed KEGG analysis of long non-coding RNA targets in the hypothalamus of pubertal rats and reported significant enrichment for the ribosome pathway. In this current study, the ribosome pathway was also enriched, and the proteins in this pathway were less abundant post-puberty in the pituitary gland. Further, these ribosomal proteins formed a dominant cluster in the pituitary interaction network. A study in pubertal Brangus heifers observed high expression of the ribosomal protein L39 gene in the pituitary gland (Cánovas et al., 2014). Further, ribosomal DA proteins in the interaction

network, such as RPS5, RPS3 and RPLP1, were also listed in the networks constructed from endometrial gene expression of low and high fertility heifers (Killeen et al., 2014). Recently, miR-503-3p was proposed as a new repressor of the initiation of puberty in female mice (Tong et al., 2018). The stable overexpression of miR-503-3p in the GT1-7 cell line can influence ribosome biogenesis pathways and result in down-regulation of puberty-related genes (Tong et al., 2018). The ribosomal protein RPL22 inhibited the expression of Lin28B, a puberty-related gene (Ong et al., 2009; Rao et al., 2012). Another ribosomal protein, namely RPS7, also regulated the PI3K and MAPK signalling pathways that are involved in puberty (Bliss et al., 2010; Acosta-Martínez, 2011; Wang et al., 2013; Nelson et al., 2017; Pal et al., 2017). From the pituitary protein-protein interaction network, there were seven ribosomal proteins that interacted with UBB protein. Disruption in the UBB gene in both male and female mice resulted in infertile animals (Ryu et al., 2008). These collective evidences suggested proteins in the ribosomal pathway are relevant for pubertal development in mammals. Further work could target the DA ribosomal proteins and their targets in the pituitary gland to investigate the molecular mechanisms of puberty in *Bos indicus*.

CONCLUSION

In summary, The study confirmed the poor global correlation between mRNAs and proteins, suggesting the need for multiple omics analyses for interpreting complex biological systems. A total of 275 and 186 DA proteins, between pre-pubertal and post-pubertal Brahman heifers in the hypothalamus and pituitary gland were identified, respectively. These DA proteins may regulate puberty onset directly or indirectly. DA proteins in the hypothalamus were mainly associated with metabolic pathways, energy metabolism, and nervous system development. In agreement with previous findings of complex feedback effects on GnRH release, the KEGG pathway analysis indicated that these DA proteins are involved in pathways that convey both inhibitory and excitatory inputs for hypothalamic neurons. In the pituitary, the abundance of proteins involved in protein digestion and absorption, focal adhesion, and ECM-receptor interaction were significantly increased post-puberty. The decreased abundance of ribosomal proteins post-puberty can be interpreted in the context of these genes' inhibitory input to puberty-related genes. Proteins related to energy production and amino acid biosynthesis may have a crucial role in the neuroendocrine regulation of the pubertal process, meriting further investigation.

DATA AVAILABILITY STATEMENT

The proteomic datasets for the hypothalamus and pituitary gland have been submitted into Proteome Xchange Consortium *via* PRIDE partner repository with identifier PXD009620 and PXD009619, respectively.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Ethics Committee of The University of Queensland ethics board (Animal Ethics Approval Number QAAFI/279/12).

AUTHOR CONTRIBUTIONS

LN designed the study, performed all experiments and statistical analyses, prepared figures and wrote the manuscript. LL contributed to **Figure 2**. MF designed the study and collected samples. All authors commented on the manuscript, read and approved the final manuscript.

FUNDING

This work was supported by the start-up grant to Stephen Moore from The University of Queensland, which covered the initial animal trial.

ACKNOWLEDGMENTS

The authors acknowledge the contributions to the field experiment and sample collection of Gry Boe-Hansen, Laercio R. Porto-Neto, Lisa Kidd, and Joao Paulo A. do Rego. The authors would also like to thank Amanda Nouwens for running LC-ESI-/MS-MS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.935433/full#supplementary-material

REFERENCES

Acosta-Martínez, M. (2011). PI3K: An Attractive Candidate for the Central Integration of Metabolism and Reproduction. *Front. Endocrin.* 2, 110. doi:10.3389/fendo.2011.00110

Anderson, L., and Seilhamer, J. (1997). A Comparison of Selected mRNA and Protein Abundances in Human Liver. *Electrophoresis* 18 (3-4), 533–537. doi:10.1002/elps.1150180333

Bai, Y., Wang, S., Zhong, H., Yang, Q., Zhang, F., Zhuang, Z., et al. (2015). Integrative Analyses Reveal Transcriptome-Proteome Correlation in Biological Pathways and Secondary Metabolism Clusters in A. flavus in Response to Temperature. *Sci. Rep.* 5, 14582. doi:10.1038/srep14582

Ballester, M., Molist, J., Lopez-Bejar, M., Sánchez, A., Santaló, J., M. Folch, J., et al. (2004). Disruption of the Mouse Phospholipase C-B1 Gene in a β-lactoglobulin Transgenic Line Affects hourbility, Growth, and Fertility in Mice. *Gene* 341, 279–289. doi:10.1016/j.gene.2004.07.007

Bangstad, H.-J., Dahl-Jørgensen, K., Kjaexsgaard, P., Mevold, K., and Hanssen, K. (19931992). Urinary Albumin Excretion Rate and Puberty in Non-diabetic Children and Adolescents. *Acta Paediatr. (Oslo, Nor.* 82 (10), 857–862. doi:10.1111/j.1651-2227.1993.tb17628.x

Bliss, S. P., Navratil, A. M., Xie, J., and Roberson, M. S. (2010). GnRH Signaling, the Gonadotrope and Endocrine Control of Fertility. *Front. Neuroendocrinol.* 31 (3), 322–340. doi:10.1016/j.yfrne.2010.04.002

Bohm, D., Schwegler, H., Kotthaus, L., Nayernia, K., Rickmann, M., Kohler, M., et al. (2002). Disruption of PLC-$beta;1-Mediated Signal Transduction in Mutant Mice Causes Age-dependent Hippocampal Mossy Fiber Sprouting and Neurodegeneration. *Mol. Cell. Neurosci.* 21 (4), 584–601. doi:10.1006/mcne.2002.1199

Bond, J. F., Robinson, G. S., and Farmer, S. R. (1984). Differential Expression of Two Neural Cell-specific Beta-Tubulin mRNAs during Rat Brain Development. *Mol. Cell. Biol.* 4 (7), 1313–1319. doi:10.1128/mcb.4.7.1313-1319.1984

Cánovas, 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 (7), e102551. doi:10.1371/journal.pone.0102551

Carvalhais, V., França, A., Pier, G. B., Vilanova, M., Cerca, N., and Vitorino, R. (2015). Comparative Proteomic and Transcriptomic Profile of Staphylococcus Epidermidis Biofilms Grown in Glucose-Enriched Medium. *Talanta* 132, 705–712. doi:10.1016/j.talanta.2014.10.012

Chen, G., Gharib, T. G., Huang, C.-C., Taylor, J. M. G., Misek, D. E., Kardia, S. L. R., et al. (2002). Discordant Protein and mRNA Expression in Lung Adenocarcinomas. *Mol. Cell. Proteomics* 1 (4), 304–313. doi:10.1074/mcp.M200008-MCP200

Chenoweth, P. (1994). Aspects of Reproduction in Female *Bos indicus* Cattle: a Review. *Aust. Vet. J.* 71 (12), 422–426. doi:10.1111/j.1751-0813.1994.tb00961.x

Clapper, J. A., Snyder, J. L., Roberts, A. J., Hamernik, D. L., and Moss, G. E. (1998). Estradiol Increases Relative Amounts of Insulin-like Growth Factor Binding Protein (IGFBP)-3 in Serum and Expression of IGFBP-2 in Anterior Pituitaries of Ewes1. *Biol. Reprod.* 59 (1), 124–130. doi:10.1095/biolreprod59.1.124

Cole, T. J., Salem, S. I., Hafez, A. S., Galal, O. M., and Massoud, A. (1982). Plasma Albumin, Parasitic Infection and Pubertal Development in Egyptian Boys. *Trans. R. Soc. Trop. Med. Hyg.* 76 (1), 17–20. doi:10.1016/0035-9203(82)90007-4

Cousminer, D. L., Berry, D. J., Timpson, N. J., Ang, W., Thiering, E., Byrne, E. M., et al. (2013). Genome-wide Association and Longitudinal Analyses Reveal Genetic Loci Linking Pubertal Height Growth, Pubertal Timing and Childhood Adiposity. *Hum. Mol. Genet.* 22 (13), 2735–2747. doi:10.1093/hmg/ddt104

Crawford, J., and McNeilly, A. (2002). Co-localisation of Gonadotrophins and Granins in Gonadotrophs at Different Stages of the Oestrous Cycle in Sheep. *J. Endocrinol.* 174 (2), 179–194. doi:10.1677/joe.0.1740179

Dai, W., Chen, Q., Wang, Q., White, R. R., Liu, J., and Liu, H. (2017). Complementary Transcriptomic and Proteomic Analyses Reveal Regulatory Mechanisms of Milk Protein Production in Dairy Cows Consuming Different Forages. *Sci. Rep.* 7, 44234. doi:10.1038/srep44234

Day, M. L., and Anderson, L. H. (1998). Current Concepts on the Control of Puberty in Cattle. *J. Animal Sci.* 76 (Suppl. l_3), 1–15. doi:10.2527/1998.76suppl_31x

Day, M. L., Imakawa, K., Garcia-Winder, M., Zalesky, D. D., Schanbacher, B. D., Kittok, R. J., et al. (1984). Endocrine Mechanisms of Puberty in Heifers: Estradiol Negative Feedback Regulation of Luteinizing Hormone Secretion 1. *Biol. Reprod.* 31 (2), 332–341. doi:10.1095/biolreprod31.2.332

Day, M. L., Imakawa, K., Wolfe, P. L., Kittok, R. J., and Kinder, J. E. (1987). Endocrine Mechanisms of Puberty in Heifers. Role of Hypothalamo-Pituitary Estradiol Receptors in the Negative Feedback of Estradiol on Luteinizing Hormone Secretion1. *Biol. Reprod.* 37 (5), 1054–1065. doi:10.1095/biolreprod37.5.1054

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. (2009). Global Signatures of Protein and mRNA Expression Levels. *Mol. Biosyst.* 5 (12), 1512–1526. doi:10.1039/b908315d

DeAtley, K. L., Colgrave, M. L., Cánovas, A., Wijffels, G., Ashley, R. L., Silver, G. A., et al. (2018). Neuropeptidome of the Hypothalamus and Pituitary Gland of Indicine × Taurine Heifers: Evidence of Differential Neuropeptide Processing in the Pituitary Gland before and after Puberty. *J. Proteome Res.* 17 (5), 1852–1865. doi:10.1021/acs.jproteome.7b00875

Desjardins, C., and Hafs, H. D. (1968). Levels of Pituitary FSH and LH in Heifers from Birth through Puberty. *J. animal Sci.* 27 (2), 472–477. doi:10.2527/jas1968.272472x

Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., et al. (2016). Gene-specific Correlation of RNA and Protein Levels in Human Cells and Tissues. *Mol. Syst. Biol.* 12 (10), 883. doi:10.15252/msb.20167144

Filis, P., Kind, P. C., and Spears, N. (2013). Implantation Failure in Mice with a Disruption in Phospholipase C Beta 1 Gene: Lack of Embryonic Attachment, Aberrant Steroid Hormone Signalling and Defective Endocannabinoid Metabolism. *Mol. Hum. Reprod.* 19 (5), 290–301. doi:10.1093/molehr/gas067

Forde, N., Mehta, J. P., Minten, M., Crowe, M. A., Roche, J. F., Spencer, T. E., et al. (2012). Effects of Low Progesterone on the Endometrial Transcriptome in Cattle1. *Biol. Reprod.* 87 (5), 124. doi:10.1095/biolreprod.112.103424

Fortes, M. R. S., Kemper, K., Sasazaki, S., Reverter, A., Pryce, J. E., Barendse, W., et al. (2013a). Evidence for Pleiotropism and Recent Selection in thePLAG1region in Australian Beef Cattle. *Anim. Genet.* 44 (6), 636–647. doi:10.1111/age.12075

Fortes, M. R. S., Li, Y., Collis, E., Zhang, Y., and Hawken, R. J. (2013b). The IGF1 Pathway Genes and Their Association with Age of Puberty in Cattle. *Anim. Genet.* 44 (1), 91–95. doi:10.1111/j.1365-2052.2012.02367.x

Fortes, M. R. S., Nguyen, L. T., Weller, M. M. D. C. A., Cánovas, A., Islas-Trejo, A., Porto-Neto, L. R., et al. (2016). Transcriptome Analyses Identify Five Transcription Factors Differentially Expressed in the Hypothalamus of Post- versus Prepubertal Brahman Heifers1. *J. animal Sci.* 94 (9), 3693–3702. doi:10.2527/jas.2016-0471

Funston, R. N., Moss, G. E., and Roberts, A. J. (1995). Insulin-like Growth Factor-I (IGF-I) and IGF-Binding Proteins in Bovine Sera and Pituitaries at Different Stages of the Estrous Cycle. *Endocrinology* 136 (1), 62–68. doi:10.1210/endo.136.1.7530196

Gao, X., Ye, J., Yang, C., Luo, L., Liu, Y., Ding, J., et al. (2018). RNA-seq Analysis of lncRNA-Controlled Developmental Gene Expression during Puberty in Goat & Rat. *BMC Genet.* 19, 19. doi:10.1186/s12863-018-0608-9

Gasser, C. L., Grum, D. E., Mussard, M. L., Fluharty, F. L., Kinder, J. E., and Day, M. L. (2006). Induction of Precocious Puberty in Heifers I: Enhanced Secretion of Luteinizing Hormone1. *J. Anim. Sci.* 84 (8), 2035–2041. doi:10.2527/jas.2005-636

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., et al. (2003). Global Analysis of Protein Expression in Yeast. *Nature* 425 (6959), 737–741. doi:10.1038/nature02046

Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing Protein Abundance and mRNA Expression Levels on a Genomic Scale. *Genome Biol.* 4 (9), 117. journal article). doi:10.1186/gb-2003-4-9-117

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., et al. (2002). Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* 1 (4), 323–333. doi:10.1074/mcp.m200001-mcp200

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell. Biol.* 19 (3), 1720–1730. doi:10.1128/mcb.19.3.1720

Haider, S., and Pal, R. (2013). Integrated Analysis of Transcriptomic and Proteomic Data. *Cg* 14 (2), 91–110. doi:10.2174/1389202911314020003

Jayapal, K. P., Philp, R. J., Kok, Y.-J., Yap, M. G. S., Sherman, D. H., Griffin, T. J., et al. (2008). Uncovering Genes with Divergent mRNA-Protein Dynamics in Streptomyces Coelicolor. *PloS one* 3 (5), e2097. doi:10.1371/journal.pone.0002097

Ji, H., Yang, H., Wang, J., Guo, J., Hu, Z., Zhang, H., et al. (2014). Effect of Adenovirus-Mediated Up-Regulation of α-enolase Gene Products on Follicle-Stimulating Hormone Receptor mRNA and Luteinizing Hormone Receptor mRNA of Granular Cells from Goose F1 Follicles. *Res. Veterinary Sci.* 96 (3), 526–532. doi:10.1016/j.rvsc.2014.02.010

Jiang, H., Kuang, Y., Wu, Y., Xie, W., Simon, M. I., and Wu, D. (1997). Roles of Phospholipase C β2 in Chemoattractant-Elicited Responses. *Proc. Natl. Acad. Sci. U.S.A.* 94 (15), 7971–7975. doi:10.1073/pnas.94.15.7971

Johnston, D. J., Barwick, S. A., Fordyce, G., Holroyd, R. G., Williams, P. J., Corbet, N. J., et al. (2013). Genetics of Early and Lifetime Annual Reproductive Performance in Cows of Two Tropical Beef Genotypes in Northern Australia. *Animal Prod. Sci.* 54 (1), 1–15.

Kang, B., Jiang, D. M., Bai, L., He, H., and Ma, R. (2014). Molecular Characterisation and Expression Profiling of the ENO1 Gene in the Ovarian Follicle of the Sichuan White Goose. *Mol. Biol. Rep.* 41 (4), 1927–1935. doi:10.1007/s11033-014-3039-3

Khadra, A., and Li, Y.-X. (2006). A Model for the Pulsatile Secretion of Gonadotropin-Releasing Hormone from Synchronized Hypothalamic Neurons. *Biophysical J.* 91 (1), 74–83. doi:10.1529/biophysj.105.080630

Killeen, A. P., Morris, D. G., Kenny, D. A., Mullen, M. P., Diskin, M. G., and Waters, S. M. (2014). Global Gene Expression in Endometrium of High and Low Fertility Heifers during the Mid-luteal Phase of the Estrous Cycle. *BMC Genomics* 15, 234. doi:10.1186/1471-2164-15-234

Kim, D., Jun, K. S., Lee, S. B., Kang, N.-G., Min, D. S., Kim, Y.-H., et al. (1997). Phospholipase C Isozymes Selectively Couple to Specific Neurotransmitter Receptors. *Nature* 389 (6648), 290–293. doi:10.1038/38508

Krsmanović, L., Stojilković, S., Merelli, F., Dufour, S. M., Virmani, M. A., and Catt, K. J. (1992). Calcium Signaling and Episodic Secretion of Gonadotropin-Releasing Hormone in Hypothalamic Neurons. *Proc. Natl. Acad. Sci.* 89 (18), 8462–8466.

Kwon, W.-S., Rahman, M. S., Lee, J.-S., Yoon, S.-J., Park, Y.-J., and Pang, M.-G. (2015b). Discovery of Predictive Biomarkers for Litter Size in Boar Spermatozoa*. *Mol. Cell. Proteomics* 14 (5), 1230–1240. doi:10.1074/mcp.M114.045369

Kwon, W.-S., Rahman, M. S., Ryu, D.-Y., Park, Y.-J., and Pang, M.-G. (2015a). Increased Male Fertility Using Fertility-Related Biomarkers. *Sci. Rep.* 5, 15654. doi:10.1038/srep15654

Lau, L. Y., Nguyen, L. T., Reverter, A., Moore, S. S., Lynn, A., McBride-Kelly, L., et al. (2020). Gene Regulation Could Be Attributed to TCF3 and Other Key Transcription Factors in the Muscle of Pubertal Heifers. *Vet. Med. Sci.* 6 (4), 695–710. doi:10.1002/vms3.278

Lesmeister, J. L., Burfening, P. J., and Blackwell, R. L. (1973). Date of First Calving in Beef Cows and Subsequent Calf Production. *J. Animal Sci.* 36 (1), 1–6. doi:10.2527/jas1973.3611

Maier, T., Güell, M., and Serrano, L. (2009). Correlation of mRNA and Protein in Complex Biological Samples. *FEBS Lett.* 583 (24), 3966–3973. doi:10.1016/j.febslet.2009.10.036

Meng, Q., Wang, K., Liu, X., Zhou, H., Xu, L., Wang, Z., et al. (2017). Identification of Growth Trait Related Genes in a Yorkshire Purebred Pig Population by Genome Wide Association Studies. *Asian-Australas J. Anim. Sci.* 30 (4), 462–469. doi:10.5713/ajas.16.0548

Michels, K. M., Lee, W. H., Seltzer, A., Saavedra, J. M., and Bondy, C. A. (1993). Up-regulation of Pituitary [125I]insulin-like Growth Factor-I (IGF-I) Binding and IGF Binding Protein-2 and IGF-I Gene Expression by Estrogen. *Endocrinology* 132 (1), 23–29. doi:10.1210/endo.132.1.7678216

Nelson, V. L. B., Negrón, A. L., Reid, I., Thomas, J. A., Yang, L., Lin, R. Z., et al. (2017). Loss of PI3K P110αin the Adipose Tissue Results in Infertility and Delayed Puberty Onset in Male Mice. *BioMed Res. Int.* 2017, 1–13. doi:10.1155/2017/3756089

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Anderson, S. T., Islas-Trejo, A., et al. (2018a). STAT6, PBX2, and PBRM1 Emerge as Predicted Regulators of 452 Differentially Expressed Genes Associated with Puberty in Brahman Heifers. *Front. Genet.* 9 (87). Original Research). doi:10.3389/fgene.2018.00087

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Islas-Trejo, A., Porto-Neto, L. R., et al. (2017). Global Differential Gene Expression in the Pituitary Gland and the Ovaries of Pre- and Postpubertal Brahman Heifers. *J. animal Sci.* 95 (2), 599–615. doi:10.2527/jas.2016.092110.2527/jas2016.0921

Nguyen, L. T., Zacchi, L. F., Schulz, B. L., Moore, S. S., and Fortes, M. R. S. (2018b). Adipose Tissue Proteomic Analyses to Study Puberty in Brahman Heifers. *J. animal Sci.* 96, 2392–2398. doi:10.1093/jas/sky128

Nguyen, L. T., Zacchi, L. F., Schulz, B. L., Moore, S. S., and Fortes, M. R. S. (2018c). *Liver Proteomics to Study the Onset of Puberty in Brahman Heifers Proceedings of the World Congress on Genetics Applied to Livestock Production No. 11*. Auckland, New Zealand, 256.

Nie, L., Wu, G., and Zhang, W. (2006). Correlation between mRNA and Protein Abundance in Desulfovibrio Vulgaris: A Multiple Regression to Identify Sources of Variations. *Biochem. Biophysical Res. Commun.* 339 (2), 603–610. doi:10.1016/j.bbrc.2005.11.055

Ojeda, S. R., Lomniczi, A., Mastronardi, C., Heger, S., Roth, C., Parent, A.-S., et al. (2006). Minireview: the Neuroendocrine Regulation of Puberty: Is the Time Ripe for a Systems Biology Approach? *Endocrinology* 147 (3), 1166–1174. doi:10.1210/en.2005-1136

Ong, K. K., Elks, C. E., Li, S., Zhao, J. H., Luan, J. a., Andersen, L. B., et al. (2009). Genetic Variation in LIN28B Is Associated with the Timing of Puberty. *Nat. Genet.* 41, 729–733. doi:10.1038/ng.382

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A. C., et al. (2017). Construction of Developmental Lineage Relationships in the Mouse Mammary Gland by Single-Cell RNA Profiling. *Nat. Commun.* 8 (1), 1627. doi:10.1038/s41467-017-01560-x

Pancholi, V. (2001). Multifunctional α-enolase: its Role in Diseases. *CMLS, Cell. Mol. Life Sci.* 58 (7), 902–920. doi:10.1007/pl00000910

Payne, S. H. (2015). The Utility of Protein and mRNA Correlation. *Trends Biochem. Sci.* 40 (1), 1–3. doi:10.1016/j.tibs.2014.10.010

Rao, S., Lee, S.-Y., Gutierrez, A., Perrigoue, J., Thapa, R. J., Tu, Z., et al. (2012). Inactivation of Ribosomal Protein L22 Promotes Transformation by Induction of the Stemness Factor, Lin28B. *Blood* 120 (18), 3764–3773. doi:10.1182/blood-2012-03-415349

Rempel, L. A., and Clapper, J. A. (2002). Administration of Estradiol-17β Increases Anterior Pituitary IGF-I and Relative Amounts of Serum and Anterior Pituitary IGF-Binding Proteins in Barrows. *J. animal Sci.* 80 (1), 214–224. doi:10.2527/2002.801214x

Roberts, A. J., Funston, R. N., and Moss, G. E. (2001). Insulin-like Growth Factor Binding Proteins in the Bovine Anterior Pituitary. *Endo* 14 (3), 399–406. doi:10.1385/endo:14:3:399

Rogers, L. C., de Boer, I., Junier, M.-P., and Ojeda, S. R. (1993). Estradiol Increases Neural-specific Class II-β-Tubulin mRNA Levels in the Developing Female Hypothalamus by Regulating mRNA Stability. *Mol. Cell. Neurosci.* 4 (5), 424–431. doi:10.1006/mcne.1993.1053

Romaniello, R., Arrigoni, F., Bassi, M. T., and Borgatti, R. (2015). Mutations in α- and β-tubulin Encoding Genes: Implications in Brain Malformations. *Brain Dev.* 37 (3), 273–280. doi:10.1016/j.braindev.2014.06.002

Ryu, K.-Y., Garza, J. C., Lu, X.-Y., Barsh, G. S., and Kopito, R. R. (2008). Hypothalamic Neurodegeneration and Adult-Onset Obesity in Mice Lacking the Ubb Polyubiquitin Gene. *Proc. Natl. Acad. Sci. U.S.A.* 105 (10), 4016–4021. doi:10.1073/pnas.0800096105

Schillo, K. K., Hall, J. B., and Hileman, S. M. (1992). Effects of Nutrition and Season on the Onset of Puberty in the Beef Heifer. *J. Anim. Sci.* 70 (12), 3994–4005. doi:10.2527/1992.70123994x

Snelling, W. M., Allan, M. F., Keele, J. W., Kuehn, L. A., McDaneld, T., Smith, T. P. L., et al. (2010). Genome-wide Association Study of Growth in Crossbred Beef Cattle12. *J. animal Sci.* 88 (3), 837–848. doi:10.2527/jas.2009-2257

Tahir, M. S., Nguyen, L. T., Schulz, B. L., Boe-Hansen, G. A., Thomas, M. G., Moore, S. S., et al. (2019). Proteomics Recapitulates Ovarian Proteins Relevant to Puberty and Fertility in Brahman Heifers (*Bos indicus* L.). *Genes.* 10 (11), 923. doi:10.3390/genes10110923

Tan, N. Y., Bailey, U.-M., Jamaluddin, M. F., Mahmud, S. H. B., Raman, S. C., and Schulz, B. L. (2014). Sequence-based Protein Stabilization in the Absence of Glycosylation. *Nat. Commun.* 5, 3099. (Article). doi:10.1038/ncomms4099

Temple, J. L., and Wray, S. (2005). Bovine Serum Albumin-Estrogen Compounds Differentially Alter Gonadotropin-Releasing Hormone-1 Neuronal Activity. *Endocrinology* 146 (2), 558–563. doi:10.1210/en.2004-1117

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., et al. (2004). Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Mol. Cell. Proteomics* 3 (10), 960–969. doi:10.1074/mcp.M400055-MCP200

Tong, L., Wang, M., Zhou, Y., Li, Y., Chen, L., Xu, F., et al. (2018). MiR-505-3p, a New Repressor of Puberty Onset in Female Mice. *bioRxiv*. doi:10.1101/271718

Vargas, C. A., Elzo, M. A., Chase, C. C., Jr., Chenoweth, P. J., and Olson, T. A. (1998). Estimation of Genetic Parameters for Scrotal Circumference, Age at Puberty in Heifers, and Hip Height in Brahman Cattle. *J. animal Sci.* 76 (10), 2536–2541. doi:10.2527/1998.76102536x

Vogel, C., and Marcotte, E. M. (2012). Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses. *Nat. Rev. Genet.* 13 (4), 227–232. doi:10.1038/nrg3185

Wang, Y., Ji, T., Nelson, A. D., Glanowska, K., Murphy, G. G., Jenkins, P. M., et al. (2018). Critical Roles of αII Spectrin in Brain Development and Epileptic Encephalopathy. *J. Clin. Investig.* 128 (2), 760–773. doi:10.1172/JCI95743

Wang, Z., Hou, J., Lu, L., Qi, Z., Sun, J., Gao, W., et al. (2013). Small Ribosomal Protein Subunit S7 Suppresses Ovarian Tumorigenesis through Regulation of the PI3K/AKT and MAPK Pathways. *PloS one* 8 (11), e79117. doi:10.1371/journal.pone.0079117

Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., et al. (2003). Protein Pathway and Complex Clustering of Correlated mRNA and Protein Expression Analyses in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 100 (6), 3107–3112. doi:10.1073/pnas.0634629100

Xie, W., Samoriski, G. M., McLaughlin, J. P., Romoser, V. A., Smrcka, A., Hinkle, P. M., et al. (1999). Genetic Alteration of Phospholipase C β3 Expression Modulates Behavioral and Cellular Responses to μ Opioids. *Proc. Natl. Acad. Sci. U.S.A.* 96 (18), 10385–10390. doi:10.1073/pnas.96.18.10385

Xu, Y., Bailey, U.-M., and Schulz, B. L. (2015). Automated Measurement of Site-specific N -glycosylation Occupancy with SWATH-MS. *Proteomics* 15 (13), 2177–2186. doi:10.1002/pmic.201400465

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fvets.2022.909039

Genome-Wide Detection of Copy Number Variations and Evaluation of Candidate Copy Number Polymorphism Genes Associated With Complex Traits of Pigs

Chunlei Zhang[1], *Jing Zhao*[1], *Yanli Guo*[1], *Qinglei Xu*[1], *Mingzheng Liu*[1], *Meng Cheng*[1], *Xiaohuan Chao*[1], *Allan P. Schinckel*[2] *and Bo Zhou*[1]*

[1] College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, China, [2] Department of Animal Sciences, Purdue University, West Lafayette, IN, United States

OPEN ACCESS

Edited by:
Nuno Carolino,
Instituto Nacional Investigaciao Agraria e Veterinaria (INIAV), Portugal

Reviewed by:
Wilson Nandolo,
Lilongwe University of Agriculture and Natural Resources, Malawi
Shabana Naz,
Government College University, Faisalabad, Pakistan

***Correspondence:**
Bo Zhou
zhoubo@njau.edu.cn

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

Received: 31 March 2022
Accepted: 09 June 2022
Published: 30 June 2022

Citation:
Zhang C, Zhao J, Guo Y, Xu Q, Liu M, Cheng M, Chao X, Schinckel AP and Zhou B (2022) Genome-Wide Detection of Copy Number Variations and Evaluation of Candidate Copy Number Polymorphism Genes Associated With Complex Traits of Pigs. Front. Vet. Sci. 9:909039. doi: 10.3389/fvets.2022.909039

Copy number variation (CNV) has been considered to be an important source of genetic variation for important phenotypic traits of livestock. In this study, we performed whole-genome CNV detection on Suhuai (SH) (n = 23), Chinese Min Zhu (MZ) (n = 11), and Large White (LW) (n = 12) pigs based on next-generation sequencing data. The copy number variation regions (CNVRs) were annotated and analyzed, and 10,885, 10,836, and 10,917 CNVRs were detected in LW, MZ, and SH pigs, respectively. Some CNVRs have been randomly selected for verification of the variation type by real-time PCR. We found that SH and LW pigs are closely related, while MZ pigs are distantly related to the SH and LW pigs by CNVR-based genetic structure, PCA, V_{ST}, and QTL analyses. A total of 14 known genes annotated in CNVRs were unique for LW pigs. Among them, the cyclin T2 (CCNT2) is involved in cell proliferation and the cell cycle. The FA Complementation Group M (FANCM) is involved in defective DNA repair and reproductive cell development. Ten known genes annotated in 47 CNVRs were unique for MZ pigs. The genes included glycerol-3-phosphate acyltransferase 3 (GPAT3) is involved in fat synthesis and is essential to forming the glycerol triphosphate. Glutathione S-transferase mu 4 (GSTM4) gene plays an important role in detoxification. Eleven known genes annotated in 23 CNVRs were unique for SH pigs. Neuroligin 4 X-linked (NLGN4X) and Neuroligin 4 Y-linked (NLGN4Y) are involved with nerve disorders and nerve signal transmission. IgLON family member 5 (IGLON5) is related to autoimmunity and neural activities. The unique characteristics of LW, MZ, and SH pigs are related to these genes with CNV polymorphisms. These findings provide important information for the identification of candidate genes in the molecular breeding of pigs.

Keywords: evolution, genetic structure analysis, economic traits, livestock, crossbreeding

INTRODUCTION

Copy number variation (CNV) was discovered in 1936 by Bridges in drosophila (1). The duplication of a segment of the drosophila *Bar* gene caused failure in the formation of normal compound eyes. The definition of CNV is constantly being refined with the additional research. Redon et al. (2) defined CNV as a DNA fragment whose copy number has changed in contrast to the reference genome, and the size from 1 kb to several Mb. According to its structural characteristics, a CNV can be classified as copy number gain or copy number loss. When both copy number gain and loss occur, it is called both type. The CNV mainly affects gene expression through gene dose-effect and gene interruption (3). When the copy number variation region (CNVR) contains dose-sensitive genes, the gene expression level changes with the copy number or the CNV in the coding region influences the gene function and leads to gene disruption and loss of coding ability.

A previous study detected 3,131 CNVRs in Chinese and European pigs. There were 129 and 147 unique CNVRs in Chinese pigs and European pigs, respectively (4). According to the functional enrichment analysis, the genes containing unique CNVRs in Chinese pig breeds are associated with disease resistance and high fertility, while the genes containing unique CNVRs in European pig breeds are closely related to muscle development (4). These results are consistent with the characteristics of Chinese and European pig breeds. A comprehensive CNV study on 98 Xiang pigs and 22 Kele pigs detected 172 CNVRs in 660 annotated genes, which are enriched in sensory, cognitive, reproductive, and ATP synthesis functions (5). These functions are well-matched with the living environment and breed characteristics of Xiang pigs and Kele pigs. In particular, the genes of propagation-related CNVRs have obvious contact with the number of piglets in the Xiang pigs. In addition, studies on the Italian white pig (6), Taihu pig (7), and Bama pig (8) also found a correlation between the breed characteristics and the functions of genes annotated in CNVRs. These studies indicate that the functions of CNVRs are associated with the phenotypes of pigs.

Large White (LW) pigs are well known for their growth and reproductive performance. Min Zhu (MZ) pigs are distributed in northern China and have the characteristics of substantial fat deposition and excellent stress resistance. Suhuai (SH) pigs are crossbred pigs that contain 75 % LW and 25 % Chinese Huai. The Huai and MZ pig breeds originated in north China. The objective of this study is to explore the characteristics of CNV in European LW, Chinese MZ, and crossbred SH pigs at the whole-genome level.

MATERIALS AND METHODS

Samples and Data

Twenty-three SH pigs were selected from the Huaiyin pig farm in Huai'an, Jiangsu Province. A standard phenol/chloroform/isoamyl alcohol protocol was used to extract genomic DNA from pig ear tissue samples. The Illumina Hiseq2000 platform was used for whole-genome sequencing. In addition, the whole-genome sequencing data of MZ pigs ($n = 11$) and LW pigs ($n = 12$) were downloaded from the public database (https://www.ncbi.nlm.nih.gov/) (**Supplementary Table 1**). The FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to analyze the quality of the sequencing data and the parameter was as follows: fastqc -o output -t thread seqfile1..seqfileN. Where "-o" indicates the pathway of the out file, "-t" indicates the number of threads running programs, and "seqfile" indicates the input sequencing data. Then the Cutadapt (https://cutadapt.readthedocs.io/en/stable/) was used for quality filtering and reads trimming. The parameter was as follows: cutadapt -q 10,15 –quality-base = 33 -o output.fastq input.fastq. Where "-q" indicates filtering the quality of the reads, 10 and 15 represent the threshold of the 3' and 5', "–quality-base = 33" indicates the phred33 score system, and "-o" indicates the pathway of the out file. The sequencing data were integrated by MultiQC (v1.11) to meet the requirements of CNV detection (**Supplementary Figure 1**) (9). The sequences were aligned to the reference genome (Sscrofa 11.1) assembly using the Burrows-Wheeler Aligner (BWA) (v 0.7.17) (10). The overall average sequencing depth reaches 12.89 ×, up to 16.22 ×, at lowest 9.16 ×, and 46 samples' average mapping ratio reached 96.47%.

CNVR's Definition and Statistics

We used the software CNVcaller to detect CNVs and determine the CNVRs (11). All steps were conducted using the default program. First, build a reference genome database. The reference genome was based on the sliding window of the user's specified size, and the GC, repeat, and gap content of each window on the genome were counted on the genome. The command was as follows: Perl CNVReferenceDB.pl reference.fa -w 800. Where "reference.fa" is the reference genome, "-w" indicates the size of the sliding window. According to the author's suggestion, we selected a window size of 800 bp, and a step of 400 bp to generate the reference genome database. Second, the absolute copy number of each window was calculated. The BAM file (BWA comparison generation) of each sample and the number of reads in each window were analyzed. The high similarity reads (≥97%) were merged, and the low-complexity regions were removed. Based on the GC content, the correct the number of reads in each window after merging was used to calculate the absolute copy number. The command was as follows: bash Individual.Process.sh -b sample.bam -h sample -d link. Where "-b" indicates the BAM file, "-h" indicates the label of the BAM file, and "-d" indicates the link files required for correction. The third step was determination of the CNVR. The boundary of each CNVR was preliminarily determined by comprehensively considering the distribution of absolute copy number, the frequency of variation, and the significant correlation between adjacent windows (primaryCNVR). Then, the adjacent CNVRs whose copy number distribution was significantly related to the population were further merged to obtain the final CNV detection results (mergedCNVR). The command was as follows: bash CNV.Discovery.sh -l list -e exclude_list -f 0.1 -h 1 -r 0.5 -p primaryCNVR -m mergeCNVR. Where "-l" indicates the list of results files after the absolute copy number correction; "-e" indicates the samples in this list are not used for the detection

of CNVR. "-f" and "-h" represent the difference between the individual's absolute copy number and the reference absolute copy number in frequency and quantity, which greater than the setting value is considered a candidate CNV window; "-r" indicates the correlation coefficient of the absolute copy number of the adjacent candidate CNV window (no overlap), which greater than the setting value will be merged; and "-p" and "-m" indicate the output files primaryCNVR and mergeCNVR. A genome-wide CNVR map was drawn by RIdeogram (12).

Genetic Structure Analysis

The CNVRs detected were used to analyze the genetic structure differences among three pig breeds. We performed principal component analysis (PCA) by PLINK (v 1.90) (13). PLINK was used to convert the CNVRs file into bed format. ADMIXTURE (v 1.3.0) was used to execute population genetic structure analysis (14). We first set the ancestral population number K value between 1 and 5, then compute the Cross-Validation Error for each K values. When the Cross-Validation Error value became the least, the K value was the number of ancestors. MEGAX was used for evolutionary tree analysis to evaluate the genetic distance between the populations. By calculating the V_{ST} value (2), we analyze the genetic difference between the two groups.

$$V_{ST} = \frac{V_{total} - \frac{V_1 \times N_1 + V_2 \times N_2}{N_{total}}}{V_{total}}$$

Where V_{total} is the total variance in copy number between the two groups, V_1 and V_2 are variances in copy number of population 1 and population 2, respectively. N_1 and N_2 are the numbers of samples of population 1 and population 2, respectively. N_{total} is the total number of all the samples. We compare the genetic distance between groups by the mean V_{ST} values. All diagrams were drawn by ggplot2 (15, 16).

CNVR Annotation and Population Differences Comparison

To further study the relationship between CNVRs and the phenotypic characteristics of the population, a Venn diagram was drawn by TBTOOLS (v 1.098661) (17) to observe the differential and common CNVRs. Gene annotation and pathway enrichment were conducted for the population-specific CNVRs using g:Profiler (18) and KOBAS (19), respectively.

Group-Specific CNVR Overlapped With QTLs

QTL data were downloaded from Pig QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index). Bedtools (v 2.15.0) (20) was used to overlap the QTLs with the group-specific CNVRs, and the unique corresponding QTL area was obtained after removing the repeat value. According to the description of QTL traits, the group-specific CNVRs that affect the phenotypes of LW, MZ, and SH pigs were analyzed.

Validation of Quantitative Real-Time PCR

We randomly selected 4 CNVRs fragments to detect copy number polymorphisms by qPCR and the $2^{-\Delta Ct}$ method, ΔCt value = (Ct_{target} – $Ct_{reference}$) (21). Primers used in qPCR were designed by Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast). The highly conserved fragment of the *GCG* in pigs was selected as an internal reference gene (22). Primer sequences for CNVRs and *GCG* are shown in **Supplementary Table 2**. To ensure that the test samples were comparable to the *GCG*, we first constructed the standard curve of each CNVR after gradient dilution of DNA. Total CNVs were verified on the QuantStudio 5 real-time PCR system (ABI, USA), and PCR amplification conditions were designed according to the manufacturer's description (Vazyme, China). The PCR amplification system was completed in a 20 μL system, including the following ingredients: 10 μL SYBR master Mix, 2 μL DNA (around 5ng), 0.4 μL forward primers, and 0.4 μL reverse primers, and 7.2 μL water. The PCR conditions were as follows: first step 95° C for 30 seconds followed by 40 cycles at 95 ° C for 10 s and 60 ° C for 30 s. The CNV type detected by the above PCR method was the same as those detected by the CNVcaller. Where CNVR-9017 was the gain type in LW pigs, but the normal type in SH pigs. The CNVR-1169, CNVR-9126, and CNVR-1771 were expressed in two pig breeds as the gain type. In addition, we used the Integrative Genomics Viewer (IGV) (23) to visualize the genome of the samples, and its results were the same as qPCR (**Supplementary Figure 2**). Each CNVR fragment has 4 biological repetitions in both LW and SH pigs, and all samples were performed in triplicate.

RESULTS

CNVR Detection and Statistics

A total of 11,173 CNVRs were detected in 46 pigs (**Supplementary Table 3**). There were 10,917, 10,885, and 10,836 CNVRs detected in SH, LW, and MZ pigs, respectively. The coverage area of these CNVRs in the three populations is more than 43 million bp, which accounts for about 1.8% of the whole genome (Sscrofa 11.1) (**Supplementary Table 4**). In all samples, there were 3,457, 2390, and 5,326 cases of copy number loss, copy number gain, and both type, respectively (**Figure 1**). The length of CNVRs ranges from 1.6 to 560 kb, but 61.23% of CNVRs are 1.6 to 3 kb, and only 0.75% CNVRs are more than 30 kb (**Supplementary Figure 3**). Moreover, a total of 8,247 CNVRs were detected in <5 pigs, and 4,134 CNVRs were found in the unique individual (**Supplementary Figure 4**).

Analysis of Population Clustering

A PCA graph was developed with all the samples having been divided into three groups: SH, MZ, and LW pig breeds (**Figure 2A**). The LW and SH pigs are closer in the PCA diagram, and the individuals are arrayed tight. The MZ pigs are far from them, and the individuals are scattered.

Genetic Structure Analysis

When the ancestral population number K = 2, there are obvious differences between LW and MZ pigs, while the information of SH pigs is covered by that of LW pigs. When K = 3, the Cross-Validation Error is the smallest (**Supplementary Table 5**), and the three pig breeds are well separated (**Figure 2B**). The result

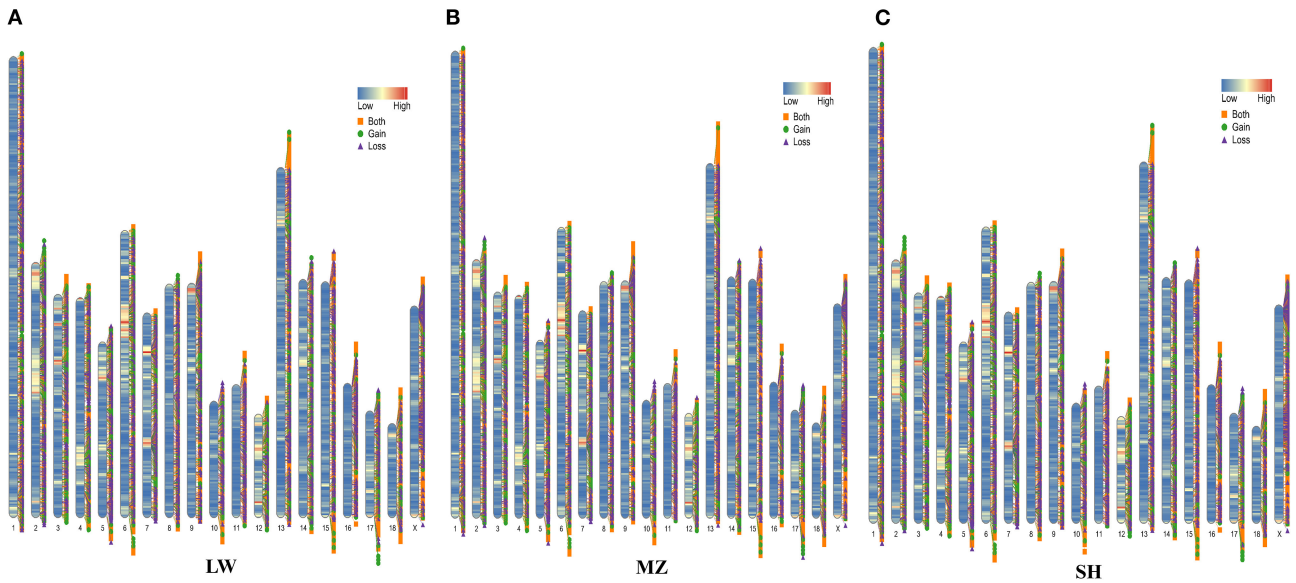


FIGURE 1 | The genome distribution of CNVRs and variation types of LW **(A)**, MZ **(B)**, and SH **(C)** pigs. The legend of "Low" to "High" indicates the gene density on the pig chromosomes. The yellow square represents the both type, the green circle represents the gain type, and the purple triangle represents the loss type.

of phylogenetic tree analysis is similar to that of PCA. Since the genetic background of the SH pig is complicated (containing 75 % Large White and 25 % Chinese Huai), the position of the SH pigs is close to the root of the tree, and the distance to LW pigs is closer than MZ pigs (**Figure 2C**). The average V_{ST} value of SH and LW pigs is just 0.111; but the average V_{ST} values are 0.234 and 0.265 in SH and MZ pigs and LW and MZ pigs, respectively (**Figure 3**). The V_{ST} analysis results are the same as the PCA analysis and genetic structural analysis. The genetic distance between SH and LW pigs is smaller than that between LW and MZ pigs.

Analysis of Shared and Group-Specific CNVR

The differences in CNVRs between pig breeds were compared through the Venn diagram (**Figure 4A**). A total of 10,671 CNVRs are shared among the three pig breeds. There are 23, 47, and 39 group-specific CNVRs in the SH, MZ, and LW pigs, respectively. A total of 140 CNVRs are common in the SH and LW pigs, while only 83 CNVRs are common in the SH and MZ pigs, and 35 CNVRs are common in the LW and MZ pigs.

Gene Research in Group-Specific CNVR

We noted the genes associated with group-specific CNVRs and discovered 35 known genes (**Table 1**) and 25 novel genes (**Supplementary Table 6**). These known genes were analyzed in the KEGG pathway.

A total of 14 known genes were annotated in 39 unique CNVRs in LW pigs. These genes regulate the metabolism of phenylalanine, histidine, and other amino acids based on the KEGG pathway (**Figure 4B**). The *CCNT2* gene is widely involved in the regulation of cell differentiation and the cell cycle. In fibroblasts of C2C12 cells, the overexpression of *CCNT2* strengthened *MyoD*-dependent transcription and promoted myogenic differentiation (24). A comprehensive study reported that the *CCNT2* gene induced the differentiation of muscle cells with the molecular partner *Pkn* (25), which may play a positive role in the meat production of LW pigs. The *FANCM* gene is involved in defective DNA repair and reproductive cell development (26). Previous studies found that the *FANCM* gene was associated with Non-obstructive Azoospermia and ovarian deficiency, which led to male/female infertility (27, 28). It may be related to the reproductive performance of the LW pigs. LW pigs are commonly mated to other maternal lines to produce crossbred commercial sows.

We annotated 10 known genes in 47 unique CNVRs in MZ pigs. These genes are enriched in "Antifolate Resistance," "Metabolic Pathways," and "Glycerolipid Metabolism" based on the KEGG pathway (**Figure 4C**). A previous study reported that the *GPAT3* gene plays an important role in lipid metabolism, which causes rapid growth and exquisite meat quality in Yunling cattle (29). The knockout of the *GPAT3* gene altered energy balance in diet-induced obesity in mice, indicating that the *GPAT3* gene plays a role in regulating energy and lipid homeostasis (30). It may be related to the fat deposition capacity of MZ pigs. The *GSTM4* and *TBC1D14* genes are considered to participate in detoxification and autophagy (31, 32). These genes are related to "Glutathione Metabolism," "Platinum Drug Resistance," and "Metabolism of Xenobiotics by Cytochrome P450" detoxification and resistance gene pathways.

We have annotated 11 known genes in the 23 unique CNVRs in SH pigs. These genes are enriched in resistance and ATP-related pathways (**Figure 4D**). Interestingly, some genes are associated with neurodevelopment. The *NLGN4X* and

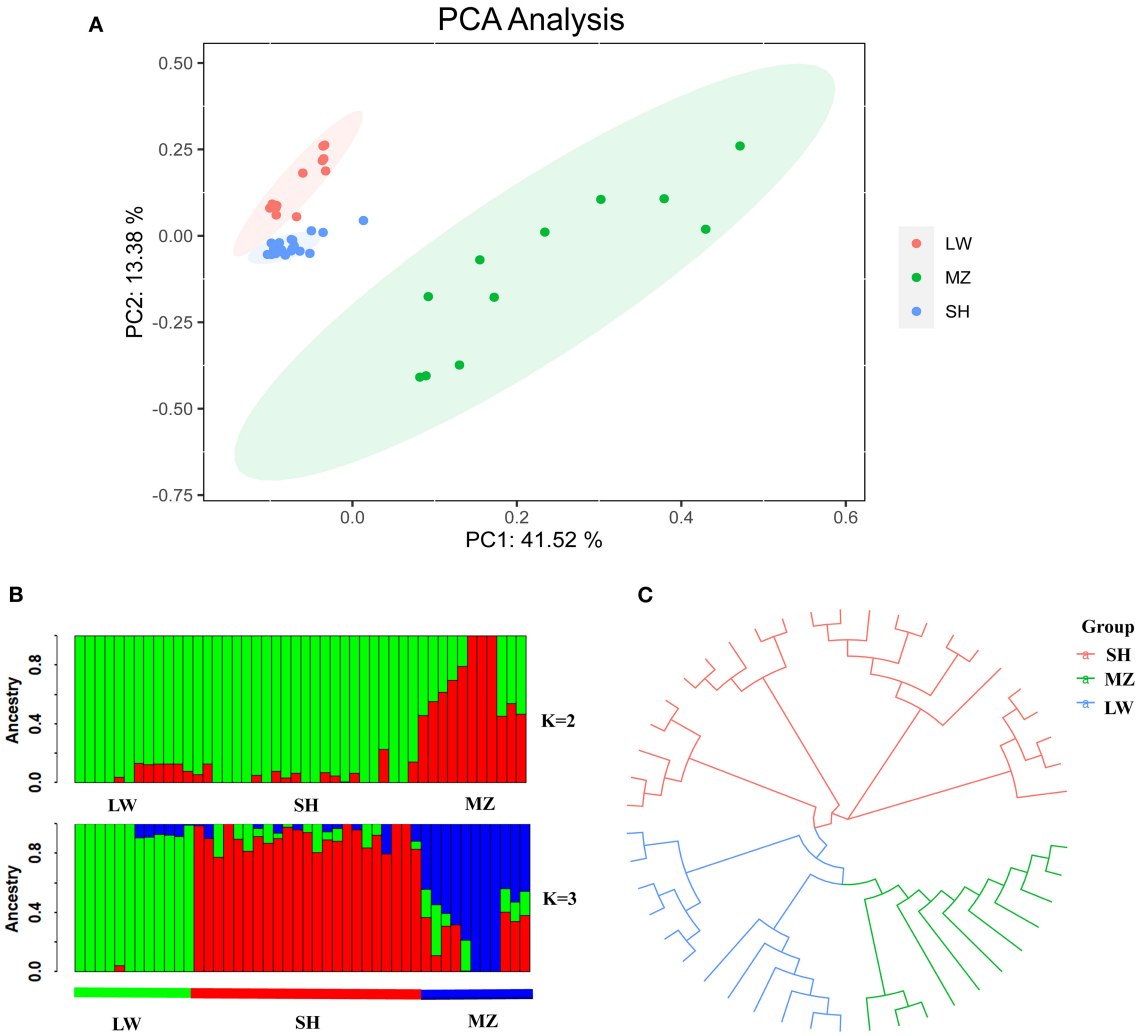


FIGURE 2 | (A) PCA plot of LW, MZ, and SH pigs. Red, green, and blue represent LW, MZ, and SH pigs, respectively. **(B)** Diagram of genetic structure analysis, the K value represents the number of the hypothetical ancestor. When the ancestral population number K = 2, there are obvious differences between LW and MZ pigs, while the information of SH pigs is covered by LW pigs, green and red represent LW and MZ pigs, respectively. When K = 3, the three pig breeds are well separated. Green, red, and blue represent LW, SH, and MZ pigs, respectively. **(C)** Evolutionary tree diagram of LW, SH, and MZ pigs. The location of SH pigs is closer to the root, moreover, the genetic distance between SH and LW pigs is less than SH and MZ pigs.

NLGN4Y genes are located on the X and Y chromosomes, respectively. Neurogenesis, neuron differentiation, and muscle development are increasingly disturbed in neuron stem cells with NLGN4X knockdown, including DLG4 and NLGN3 postsynaptic genes also have decreased expression (33). The *IGLON5* gene participates in regulating sleep and other neural activities and is also related to autoimmunity (34).

Group-Specific CNVRs Overlapped With QTLs

The group-specific CNVRs of LW, SH, and MZ pigs were mapped in the QTLs of the pigs. There are 1,139, 938, and 1,283 QTLs in the SH, LW, and MZ pigs, respectively. A Venn diagram shows that 248 QTLs overlap between the LW and SH pigs, 237 QTLs overlap between the SH and MZ pigs, and 178 QTLs overlap between the MZ and LW pigs. There are 285, 545, and 700 group-specific QTLs in the SH, LW, and MZ pigs, respectively (**Supplementary Figure 5**). A circus diagram was used to show the location of these unique QTLs (**Figure 5A**). The effects of QTLs on traits are divided into three levels, "Trait Categories," "Trait Type," and "Trait." The difference in the meat and disease resistance traits of LW, MZ, and SH pigs is more distinct (35) (**Figures 5B–D**). So QTLs for meat and health trait categories were analyzed.

In the anatomy type of the meat category, the trait cases of "muscle area and muscle fiber" and "fat to meat ratio and fat-cut percentage" are different in LW, MZ, and SH pigs. The number of muscle-related QTLs is 12.2 times that of fat-related QTLs in LW pigs (61/5). And this ratio is only 3.6 times and 7 times in MZ and SH pigs (55/15, 35/5). Interestingly, the "EnzyMeactivity"

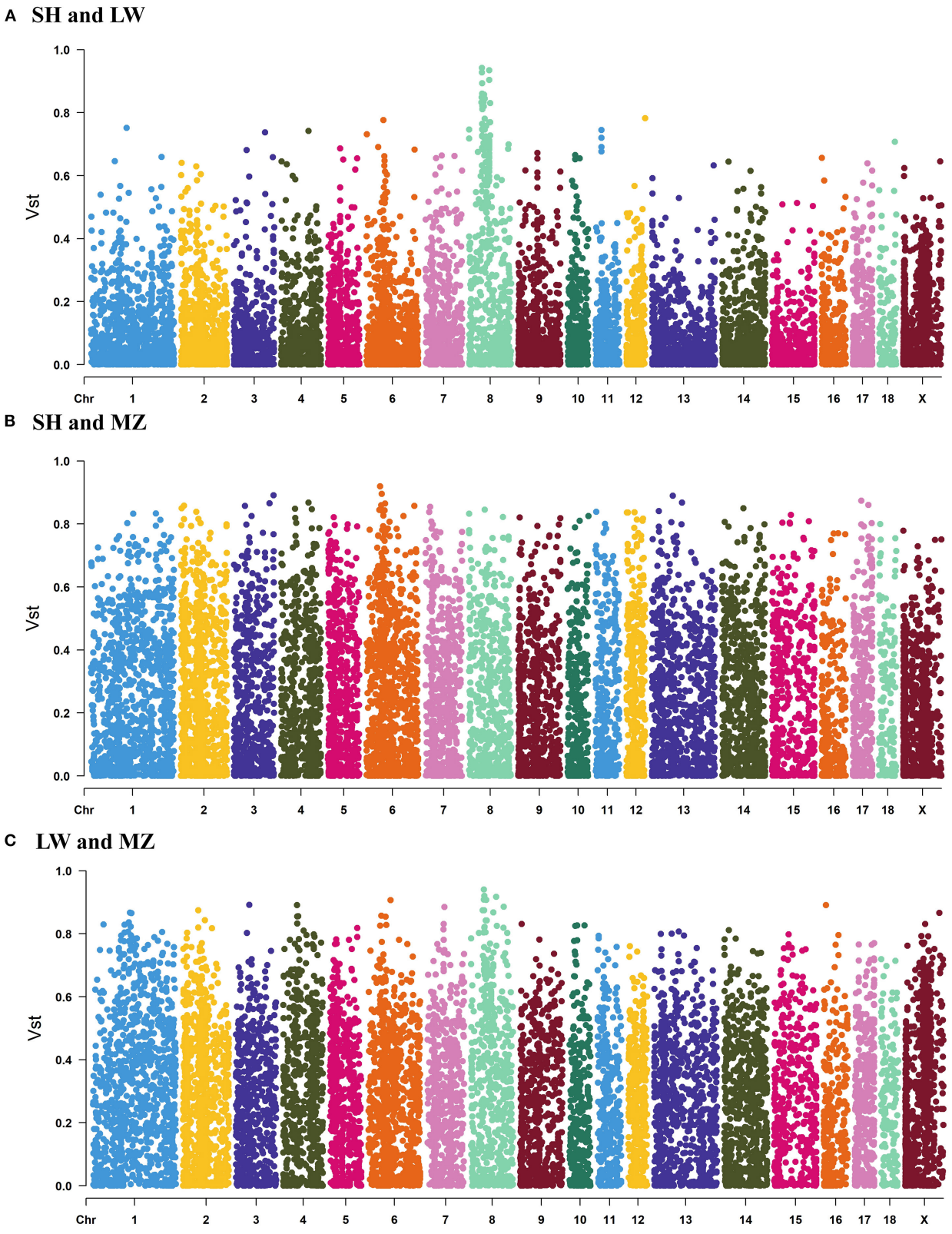


FIGURE 3 | The V_{ST} values of all the copy number variation regions (CNVRs) in SH and LW **(A)** pigs, SH and MZ **(B)** pigs, LW and MZ **(C)** pigs. The average VST value of SH and LW pigs is just 0.111; but the average VST values are 0.234 and 0.265 in SH and MZ pigs and LW and MZ pigs, respectively.

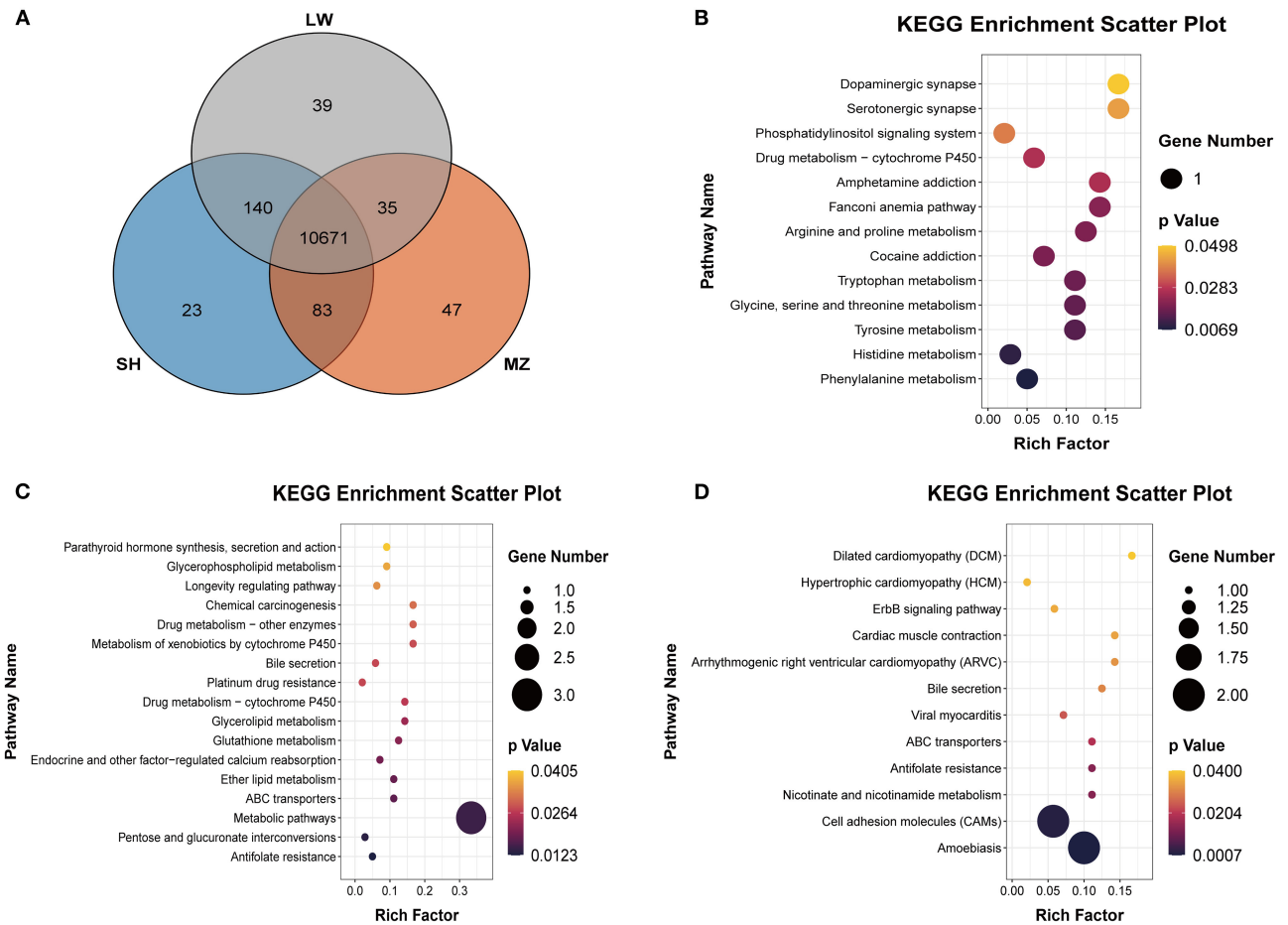


FIGURE 4 | (A) Venn diagram of the CNVRs in LW, SH, and MZ pigs. The known genes of LW **(B)**, MZ **(C)**, and SH **(D)** pigs in the group-specific CNVRs were analyzed in the KEGG pathway.

QTLs are unique to the MZ pigs. The number of total "NADPH-generation enzyme activity" and "NADP-malate dehydrogenase activity" is 12, which is related to the oxidation reaction in the organism, particularly fatty acids generation (36). In the trait category of health, the number of "Immune capacity" is huge difference among LW, MZ, and SH pigs, with a total of 24 traits, and 64 QTLs related to immune capacity in the MZ pigs, but only 6 traits, 23 QTLs, and 14 traits, 34 QTLs are in LW and SH pigs, respectively.

DISCUSSION

The role of CNV's is an increasingly discussed academic topic, and previous studies on CNV have been conducted in humans, cattle, sheep, and other species (37–40). CNVs could destroy the normal expression of genes and ultimately cause phenotypic changes mainly through dosage effects, interruption, and position effects of gene deletion and duplication (41–43). As a type of essential variation in the genome, CNV polymorphisms play key roles in species evolution, environmental adaptation, disease resistance, and disease susceptibility (44–46). However, numerous past studies have concentrated on CNV on the chromosomal DNA with little attention given to CNV of non-chromosomal DNA. Mitochondrial DNA (mtDNA) passes through maternal inheritance, which has been confirmed to be related to many traits, including respiratory and cardiovascular disease (47). As a component of ribosomes, rRNA easily becomes a substrate of homologous recombination resulting in CNV due to its repetitive sequence structure (48).

In our present study, we noticed that LW pigs have excellent meat production. Several genes containing the unique CNVRs are involved in the regulation of cell proliferation and cell cycle regulation in LW pigs. These genes have extensive participation in muscle growth and development. We also obtained the same results in the QTL analysis.

Among these genes, the *CCNT2* gene is related to cell differentiation and cell cycle, especially regulating the differentiation of muscle cells (24). Many studies have focused on the combined analysis of microRNA (miRNA) and *CCNT2*. Previous research reported that miR-15a,

TABLE 1 | Annotated genes in group-Specific CNVRs of LW, MZ, and SH pigs.

Population	Ensemble_id	Gene_name	Description
	ENSSSCG00000032786	ZC3HAV1L	Zinc finger CCCH-type containing, antiviral 1 like
	ENSSSCG00000024999	PPIP5K2	Diphosphoinositol pentakisphosphate kinase 2
	ENSSSCG00000025784	CDH4	Cadherin 4
	ENSSSCG00000015830	UNC5D	Unc-5 netrin receptor D
	ENSSSCG00000005000	FANCM	FA complementation group M
	ENSSSCG00000016141	PLEKHM3	Pleckstrin homology domain containing M3
LW	ENSSSCG00000015379	DNAH11	Dynein axonemal heavy chain 11
	ENSSSCG00000001975	PRKD1	Protein kinase D1
	ENSSSCG00000015697	CCNT2	Cyclin T2
	ENSSSCG00000023215	MAOB	Monoamine oxidase B
	ENSSSCG00000012101	ANOS1	Anosmin 1
	ENSSSCG00000042659	ZSCAN5A	Zinc finger and SCAN domain containing 5A
	ENSSSCG00000042659	ZSCAN5B	Zinc finger and SCAN domain containing 5B
	ENSSSCG00000042659	ZSCAN5C	Zinc finger and SCAN domain containing 5C
	ENSSSCG00000009347	KL	Klotho
	ENSSSCG00000044340	EEA1	Early endosome antigen 1
	ENSSSCG00000027349	TBC1D14	TBC1 domain family member 14
	ENSSSCG00000009233	GPAT3	Glycerol-3-phosphate acyltransferase 3
	ENSSSCG00000007727	AUTS2	Activator of transcription and developmental regulator AUTS2
MZ	ENSSSCG00000030262	GDPD1	Glycerophosphodiester phosphodiesterase domain containing 1
	ENSSSCG00000038036	TTLL11	Tubulin tyrosine ligase like 11
	ENSSSCG00000037808	GSTM4	Glutathione S-transferase mu 4
	ENSSSCG00000021846	EFHC2	EF-hand domain containing 2
	ENSSSCG00000009497	ABCC4	ATP binding cassette subfamily C member 4
	ENSSSCG00000011040	CACNB2	Calcium voltage-gated channel auxiliary subunit beta 2
	ENSSSCG00000015435	NAMPT	Nicotinamide phosphoribosyltransferase
	ENSSSCG00000023934	KCNIP4	Potassium voltage-gated channel interacting protein 4
	ENSSSCG00000009215	ABCG2	ATP binding cassette subfamily G member 2 (Junior blood group)
	ENSSSCG00000033643	NLGN4X	Neuroligin 4 X-linked
SH	ENSSSCG00000033643	NLGN4Y	Neuroligin 4 Y-linked
	ENSSSCG00000033560	SERPINB3	Serpin family B member 3
	ENSSSCG00000033560	SERPINB4	Serpin family B member 4
	ENSSSCG00000011121	CELF2	CUGBP Elav-like family member 2
	ENSSSCG00000003227	IGLON5	IgLON family member 5
	ENSSSCG00000024674	ABL2	ABL proto-oncogene 2, non-receptor tyrosine kinase

miR-155-5p, and miR-188-5p inhibit muscle differentiation and skeletal muscle development *via* target binding *CCNT2* (49–51). Due to their great reproductive performance, in the modern pig breeding systems, LW pigs are used to produce crossbred female parents. Among these genes, *FANCM* is involved in DNA damage repair, and the mutation causes deaths of spermatogenic cells at all levels and stagnation of round spermatids, which causes male reproductive disorders, including sperm deformities, decreased motility, and decreased numbers (27). These results are interesting because these genes may be related to the reproductive performance of LW pigs.

We found *GPAT3* related to adipogenesis in unique CNVRs in MZ pigs. The promoter polymorphisms of the *GPAT3* were associated with intramuscular fat content in Laiwu pigs, and the knockout of *GPAT3* was related to insulin resistance and fatty liver in a mouse model of severe congenital generalized lipodystrophy (30, 52). The *GPAT3* accelerated the fat production capacity of MZ pigs. Understandably, the habitat of the MZ pigs is in northern China, where winter temperatures reach minus 40 degrees Celsius. Sufficient fat keeps them resistant to the cold and stores energy. Similarly, MZ pigs have good disease resistance and detoxification capabilities. *GSTM4* is a member of the glutathione sulfur transferase family and plays a key role in the detoxification of insecticides and other exogenous substances. In abamectin-resistant *tetranychus urticae*, the activity of GSTs was significantly increased (53). The QTLs mapped to the group-specific CNVRs in MZ pigs are related to fat and immunity. The genes mentioned above provide favorable conditions for the survival of MZ pigs in cold regions.

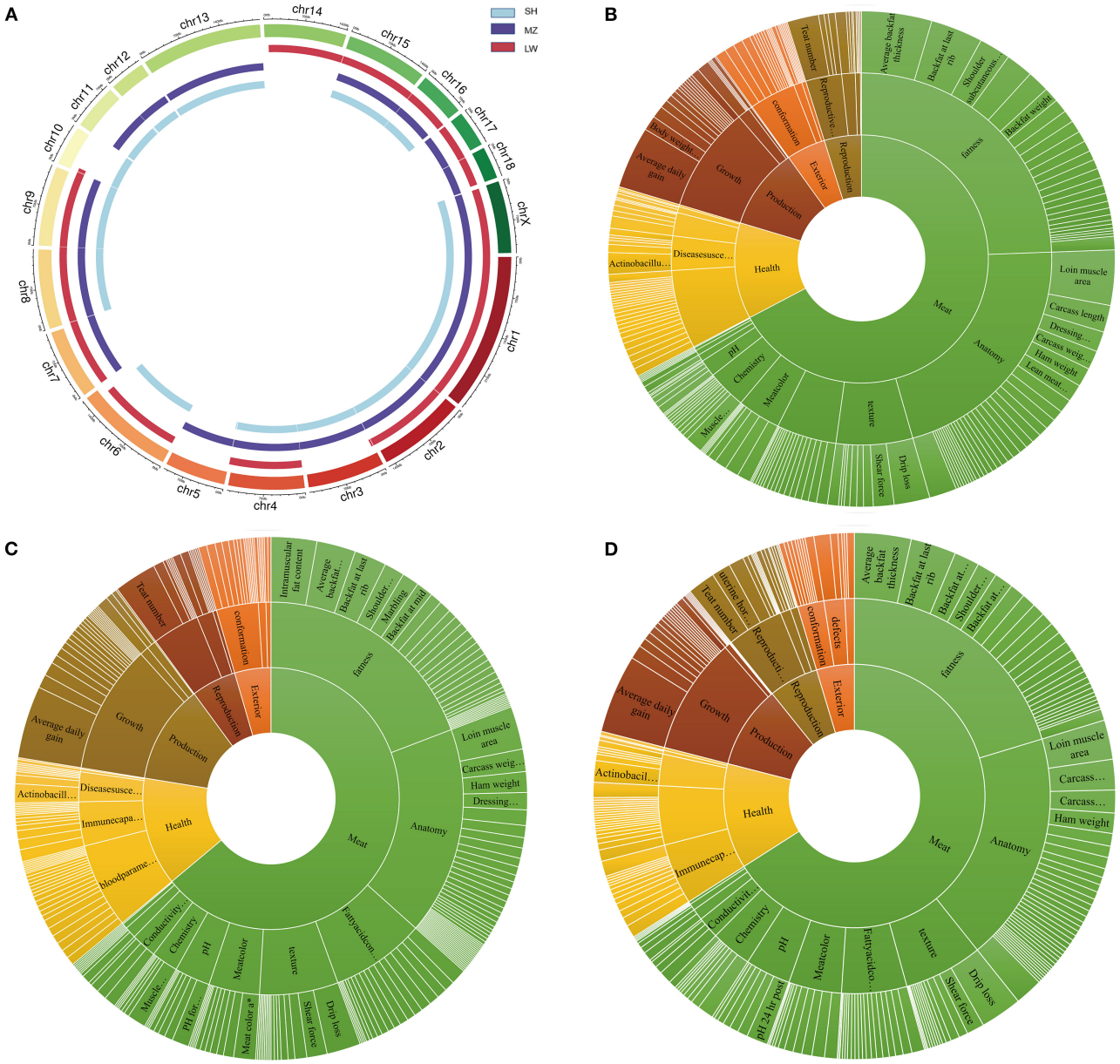


FIGURE 5 | (A) The genome distribution of the group-specific QTLs in SH, MZ, and LW pigs. **(B–D)** are the group-specific QTLs in SH, LW, and MZ pigs.

The SH pig is crossbred of Chinese and European pigs. The CNV polymorphisms of some genes were unique in SH pigs. *SERPINB3* is a homologous substance to chicken ovalbumin protein (OVA) in humans. It takes part in apoptosis and autoimmune diseases and is related to the prognosis (54). The *NAMPT* is primarily involved in redox reactions, and the signals it transmits act during various stages of cell physiology, including cell cycle and proliferation (55). It is a participant and regulator of many diseases. The results were within our expectations, including genes related to immunity and cell proliferation. What surprised us was that some genes are related to neuroprotection and neurological disorders. *NLGN4X* and *NLGN4Y*, as marker molecules of human autism, are considered to play an important role in the etiology of autism, the formation of synapses, and the transmission of information. Autism can lead to stereotypic behavior and communication difficulties in humans and is related to developmental mental disorders (56, 57). In addition, the massive accumulation of *IGLON5* antibodies has been proven to damage the cytoskeleton of hippocampal neurons, which can lead to the occurrence of autoimmune diseases and neurodegeneration (34, 58). These findings were interesting as SH pigs are more docile and more easily domesticated than LW

pigs. The neurological foundation of these behavioral differences is still unknown.

By analyzing the genetic structure of LW, MZ, and SH pigs, we found that SH and LW pigs are closely related, while MZ pigs are distantly related to pigs of the other two breeds. It indicates that LW and SH pigs have more genetic exchanges than MZ pigs, which have the same trend in PCA, evolutionary tree, V_{ST}, and the group-special CNVRs and QTLs analyses. Based on the results of genetic structural analysis, we found that the lineage of SH pigs came from LW pigs, and MZ pigs have a smaller genetic distance from SH pigs than LW pigs. This may be because the MZ pig have genetic exchanges with the LW pig of widespread reproduction, and the habitats of MZ and SH pigs are similar in geographical location, climate, and altitude, which have the same environmental driving forces and adaptability that make them produce the same CNV (59). Understandably, the main source of CNV was inherited from ancestors, followed by adaptation to environmental changes and other reasons that led to random mutations (60, 61).

CONCLUSION

In summary, we have performed genome-wide CNV detection on LW, MZ, and SH pigs to explore the relationship between CNVs and phenotypic characteristics of pig breeds. The functions of genes containing unique CNVRs are related to the phenotypic traits of pig breeds. From this, we have identified some candidate genes. These CNV polymorphisms provide a theoretical basis for the understanding of the relationship between phenotype and CNVs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by Experimental Animal Welfare and Ethics Committee of Nanjing Agricultural University, Nanjing, China.

AUTHOR CONTRIBUTIONS

BZ came up with the idea and revised the manuscript. CZ wrote the manuscript and performed the experiments. JZ, YG, and QX collected the samples and isolated the genomic DNA. ML, MC, and XC analyzed the data. AS and BZ reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (NO. 32172786) and the JBGS Project of Breeding Industry Revitalization in Jiangsu Province [JBGS(2021)101].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.909039/full#supplementary-material

Supplementary Figure 1 | All samples are suitable for CNV detection.

Supplementary Figure 2 | Genomic visualization of CNVR-9017, CNVR-1169, CNVR-9126, and CNVR-1771 in LW and SH pigs.

Supplementary Figure 3 | The length-frequency distribution of CNVRs. The majority of CNVRs are concentrated in 1.6-3 kb, accounting for 61.23% of the total, with only 0.75% exceeding 30 kb.

Supplementary Figure 4 | The variable frequency distribution of CNVRs. A total of 8,247 CNVRs were found in <5 individuals, and 4,134 CNVRs were found in a unique individual.

Supplementary Figure 5 | A Venn diagram shows 285, 545, and 700 group-specific QTLs in the SH, LW, and MZ pigs, respectively.

Supplementary Table 1 | The whole-genome sequencing data of MZ, LW, and SH pigs.

Supplementary Table 2 | The standard curve and primers for qPCR, and the verification results of the CNV type.

Supplementary Table 3 | A total of CNVRs were detected in 46 pigs and variations of types.

Supplementary Table 4 | The distribution of CNVRs on pig chromosomes of the pig.

Supplementary Table 5 | The Cross-Validation Error under the ancestral population number K value ranges from 1 to 5.

Supplementary Table 6 | Novel genes identified in LW, MZ, and SH pigs.

REFERENCES

1. BRIDGES CB. The bar "Gene" a duplication. *Science.* (1936) 83:210–1. doi: 10.1126/science.83.2148.210
2. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* (2006) 444:444–54. doi: 10.1038/nature05329
3. Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* (2005) 1:e49. doi: 10.1371/journal.pgen.0010049
4. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, et al. Analysis of genome-wide copy number variations in chinese indigenous and western pig breeds by 60 k snp genotyping arrays. *PLoS ONE.* (2014) 9:e106780. doi: 10.1371/journal.pone.0106780
5. Xie J, Li R, Li S, Ran X, Wang J, Jiang J, et al. Identification of copy number variations in xiang and kele pigs. *PLoS ONE.* (2016) 11:e0148565. doi: 10.1371/journal.pone.0148565
6. Schiavo G, Dolezal MA, Scotti E, Bertolini F, Calo DG, Galimberti G, et al. Copy number variants in italian large white pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Anim Genet.* (2014) 45:745–9. doi: 10.1111/age.12180
7. Wang Z, Chen Q, Liao R, Zhang Z, Zhang X, Liu X, et al. Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing. *Anim Genet.* (2017) 48:38–47. doi: 10.1111/age.12465

8. Zhang L, Huang Y, Si J, Wu Y, Wang M, Jiang Q, et al. Comprehensive inbred variation discovery in bama pigs using de novo assemblies. *Gene.* (2018) 679:81–9. doi: 10.1016/j.gene.2018.08.051
9. Ewels P, Magnusson M, Lundin S, Kaller M. Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354
10. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
11. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, et al. Cnvcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience.* (2017) 6:1–12. doi: 10.1093/gigascience/gix115
12. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, et al. Rideogram: drawing Svg graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci.* (2020) 6:e251. doi: 10.7717/peerj-cs.251
13. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation Plink: rising to the challenge of larger and richer datasets. *Gigascience.* (2015) 4:7. doi: 10.1186/s13742-015-0047-8
14. Shringarpure SS, Bustamante CD, Lange K, Alexander DH. Efficient analysis of large datasets and sex bias with admixture. *BMC Bioinformatics.* (2016) 17:218. doi: 10.1186/s12859-016-1082-x
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* (2018) 35:1547–9. doi: 10.1093/molbev/msy096
16. Ito K, Murphy D. Application of Ggplot2 to pharmacometric graphics. *CPT Pharmacometrics Syst Pharmacol.* (2013) 2:e79. doi: 10.1038/psp.2013.56
17. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. Tbtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* (2020) 13:1194–202. doi: 10.1016/j.molp.2020.06.009
18. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 Update). *Nucleic Acids Res.* (2019) 47:W191–W8. doi: 10.1093/nar/gkz369
19. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* (2011) 39:W316–22. doi: 10.1093/nar/gkr483
20. Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033
21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative Pcr and the 2(-Delta Delta C(T)) method. *Methods.* (2001) 25:402–8. doi: 10.1006/meth.2001.1262
22. Ballester M CA, Ibáñez E, Sánchez A, Folch JM. Real-time quantitative pcr-based system for determining transgene copy number in transgenic animals. *Biotechniques.* (2004) 37:3. doi: 10.2144/04374ST06
23. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (Igv): high-performance genomics data visualization and exploration. *Brief Bioinform.* (2013) 14:178–92. doi: 10.1093/bib/bbs017
24. Simone C SP, Bagella L, Pucci B, Bellan C, De Falco G, De Luca A, et al. Activation of myod-dependent transcription by cdk9/Cyclin T2. *Oncogene.* (2002 J) 21:4137–48. doi: 10.1038/sj.onc.1205493
25. Cottone G, Baldi A, Palescandolo E, Manente L, Penta R, Paggi MG, et al. Pkn is a novel partner of cyclin T2a in muscle differentiation. *J Cell Physiol.* (2006) 207:232–7. doi: 10.1002/jcp.20566
26. Kasak L, Punab M, Nagirnaja L, Grigorova M, Minajeva A, Lopes AM, et al. Bi-allelic recessive loss-of-function variants in fancm cause non-obstructive azoospermia. *Am J Hum Genet.* (2018) 103:200–12. doi: 10.1016/j.ajhg.2018.07.005
27. Yin H, Ma H, Hussain S, Zhang H, Xie X, Jiang L, et al. A homozygous fancm frameshift pathogenic variant causes male infertility. *Genet Med.* (2019) 21:62–70. doi: 10.1038/s41436-018-0015-7
28. Jaillard S, Bell K, Akloul L, Walton K, McElreavy K, Stocker WA, et al. New insights into the genetic basis of premature ovarian insufficiency: novel causative variants and candidate genes revealed by genomic sequencing. *Maturitas.* (2020) 141:9–19. doi: 10.1016/j.maturitas.2020.06.004
29. Zhang F, Hanif Q, Luo X, Jin X, Zhang J, He Z, et al. Muscle transcriptome analysis reveal candidate genes and pathways related to fat and lipid metabolism in yunling cattle. *Anim Biotechnol.* (2021) 7:1−8. doi: 10.1080/10495398.2021.2009846 [Epub ahead of print].
30. Gao M, Liu L, Wang X, Mak HY, Liu G, Yang H. Gpat3 deficiency alleviates insulin resistance and hepatic steatosis in a mouse model of severe congenital generalized lipodystrophy. *Hum Mol Genet.* (2020) 29:432–43. doi: 10.1093/hmg/ddz300
31. Lamb CA, Nuhlen S, Judith D, Frith D, Snijders AP, Behrends C, et al. Tbc1d14 regulates autophagy *via* the trapp complex and Atg9 traffic. *EMBO J.* (2016) 35:281–301. doi: 10.15252/embj.201592695
32. Denson J, Xi Z, Wu Y, Yang W, Neale G, Zhang J. Screening for inter-individual splicing differences in human Gstm4 and the discovery of a single nucleotide substitution related to the tandem skipping of two exons. *Gene.* (2006) 379:148–55. doi: 10.1016/j.gene.2006.05.012
33. Shi L, Chang X, Zhang P, Coba MP, Lu W, Wang K. The functional genetic link of nlgn4x knockdown and neurodevelopment in neural stem cells. *Hum Mol Genet.* (2013) 22:3749–60. doi: 10.1093/hmg/ddt226
34. Landa J, Gaig C, Plaguma J, Saiz A, Antonell A, Sanchez-Valle R, et al. Effects of Iglon5 antibodies on neuronal cytoskeleton: a link between autoimmunity and neurodegeneration. *Ann Neurol.* (2020) 88:1023–7. doi: 10.1002/ana.25857
35. Clapperton M, Bishop SC, Glass EJ. Innate immune traits differ between meishan and large white pigs. *Vet Immunol Immunopathol.* (2005) 104:131–44. doi: 10.1016/j.vetimm.2004.10.009
36. Belew GD, Silva J, Rito J, Tavares L, Viegas I, Teixeira J, et al. Transfer of glucose hydrogens *via* acetyl-coa, malonyl-coa, and nadph to fatty acids during de novo lipogenesis. *J Lipid Res.* (2019) 60:2050–6. doi: 10.1194/jlr.RA119000354
37. Liu M, Li B, Shi T, Huang Y, Liu GE, Lan X, et al. Copy number variation of bovine Shh gene is associated with body conformation traits in Chinese beef cattle. *J Appl Genet.* (2019) 60:199–207. doi: 10.1007/s13353-019-00496-w
38. Feng Z, Li X, Cheng J, Jiang R, Huang R, Wang D, et al. Copy number variation of the pigy gene in sheep and its association analysis with growth traits. *Animals.* (2020) 10:6888. doi: 10.3390/ani10040688
39. Locke ME, Milojevic M, Eitutis ST, Patel N, Wishart AE, Daley M, et al. Genomic copy number variation in mus musculus. *BMC Genomics.* (2015) 16:497. doi: 10.1186/s12864-015-1713-z
40. Khatri B, Kang S, Shouse S, Anthony N, Kuenzel W, Kong BC. Copy number variation study in japanese quail associated with stress related traits using whole genome re-sequencing data. *PLoS ONE.* (2019) 14:e0214543. doi: 10.1371/journal.pone.0214543
41. Vegesna R, Tomaszkiewicz M, Medvedev P, Makova KD. Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLoS Genet.* (2019) 15:e1008369. doi: 10.1371/journal.pgen.1008369
42. Iijima-Yamashita Y, Matsuo H, Yamada M, Deguchi T, Kiyokawa N, Shimada A, et al. Multiplex fusion gene testing in pediatric acute myeloid leukemia. *Pediatr Int.* (2018) 60:47–51. doi: 10.1111/ped.13451
43. Velagaleti GV B-WG, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M. et al. Position effects due to chromosome breakpoints that map approximately 900 kb upstream and approximately 13 Mb downstream of Sox9 in two patients with campomelic. *Am J Hum Genet.* (2005) 76:652–62. doi: 10.1086/429252
44. Wang H, Wang C, Yang K, Liu J, Zhang Y, Wang Y, et al. Genome wide distributions and functional characterization of copy number variations between Chinese and western pigs. *PLoS ONE.* (2015) 10:e0131522. doi: 10.1371/journal.pone.0131522
45. Fernandez AI, Barragan C, Fernandez A, Rodriguez MC, Villanueva B. Copy number variants in a highly inbred Iberian porcine strain. *Anim Genet.* (2014) 45:357–66. doi: 10.1111/age.12137
46. Revay T, Quach AT, Maignel L, Sullivan B, King WA. Copy number variations in high and low fertility breeding boars. *BMC Genomics.* (2015) 16:280. doi: 10.1186/s12864-015-1473-9
47. Foote K, Reinhold J, Yu EPK, Figg NL, Finigan A, Murphy MP, et al. Restoring mitochondrial DNA copy number preserves mitochondrial function and delays vascular aging in mice. *Aging Cell.* (2018) 17:e12773. doi: 10.1111/acel.12773
48. Porokhovnik L. Individual copy number of ribosomal genes as a factor of mental retardation and autism risk and severity. *Cells.* (2019) 8:1151. doi: 10.3390/cells8101151

49. Teng Y, Wang Y, Fu J, Cheng X, Miao S, Wang L. Cyclin T2: a novel Mir-15a target gene involved in early spermatogenesis. *FEBS Lett.* (2011) 585:2493–500. doi: 10.1016/j.febslet.2011.06.031
50. Xu S, Chang Y, Wu G, Zhang W, Man C. Potential role of Mir-155-5p in fat deposition and skeletal muscle development of chicken. *Biosci Rep.* (2020) 40. doi: 10.1042/BSR20193796
51. Wang F ZQ, Liu JZ, Kong DL. Mirna-188-5p alleviates the progression of osteosarcoma *via* target degrading Ccnt2. *Eur Rev Med Pharmacol Sci.* (2020) 24:29–35. doi: 10.26355/eurrev_202001_19892
52. Ma C, Sun Y, Wang J, Kang L, Jiang Y. Identification of a promoter polymorphism affecting Gpat3 gene expression that is likely related to intramuscular fat content in pigs. *Anim Biotechnol.* (2020) 21:1–4. doi: 10.1080/10495398.2020.1858847 [Epub ahead of print].
53. Mounsey KEPC, Arlian LG, Morgan MS, Holt DC, Currie BJ, Walton SF, et al. Increased transcription of glutathione S-Transferases in acaricide exposed scabies mites. *Parasit Vectors.* (2010) 3:43. doi: 10.1186/1756-3305-3-43
54. Riaz N, Havel JJ, Kendall SM, Makarov V, Walsh LA, Desrichard A, et al. Recurrent Serpinb3 and Serpinb4 mutations in patients who respond to Anti-Ctla4 IMMUNOTHERAPY. *Nat Genet.* (2016) 48:1327–9. doi: 10.1038/ng.3677
55. Sharif T, Martell E, Dai C, Ghassemi-Rad MS, Kennedy BE, Lee PWK, et al. Regulation of Cancer and Cancer-Related Genes Via Nad(). *Antioxid Redox Signal.* (2019) 30:906–23. doi: 10.1089/ars.2017.7478
56. Nguyen TA, Wu K, Pandey S, Lehr AW, Li Y, Bemben MA, et al. A cluster of autism-associated variants on X-Linked Nlgn4x functionally resemble Nlgn4y. *Neuron.* (2020) 106:759–68e7. doi: 10.1016/j.neuron.2020.03.008
57. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, et al. Mutations of the X-Linked genes encoding neuroligins Nlgn3 and Nlgn4 are associated with autism. *Nat Genet.* (2003) 34:27–9. doi: 10.1038/ng1136
58. Ryding M, Gamre M, Nissen MS, Nilsson AC, Okarmus J, Poulsen AAE, et al. Neurodegeneration induced by anti-Iglon5 antibodies studied in induced pluripotent stem cell-derived human neurons. *Cells*. (2021) 10:837. doi: 10.3390/cells10040837
59. Frantz LA, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of eurasian wild and domestic pig genomes. *Nat Genet.* (2015) 47:1141–8. doi: 10.1038/ng.3394
60. Hull RM, Cruz C, Jack CV, Houseley J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* (2017) 15:e2001333. doi: 10.1371/journal.pbio.2001333
61. Stalder L, Oggenfuss U, Mohd-Assaad N, Croll D. The population genetics of adaptation through copy-number variation in a fungal plant pathogen. *Mol Ecol.* (2022) 27:1633-50. doi: 10.1111/mec.16435

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fgene.2022.920641

Loss of Monoallelic Expression of *IGF2* in the Adult Liver *Via* Alternative Promoter Usage and Chromatin Reorganization

Jinsoo Ahn[1], *Joonbum Lee*[1,2], *Dong-Hwan Kim*[1], *In-Sul Hwang*[3,4], *Mi-Ryung Park*[3], *In-Cheol Cho*[5], *Seongsoo Hwang*[3] *and Kichoon Lee*[1,2]*

[1]*Functional Genomics Laboratory, Department of Animal Sciences, The Ohio State University, Columbus, OH, United States,* [2]*The Ohio State University Interdisciplinary Human Nutrition Program, The Ohio State University, Columbus, OH, United States,* [3]*Animal Biotechnology Division, National Institute of Animal Science, Rural Development Administration, Jeonbuk, South Korea,* [4]*Columbia Center for Translational Immunology, Columbia University Irving Medical Center, Columbia University, New York, NY, United States,* [5]*National Institute of Animal Science, Rural Development Administration, Jeju, South Korea*

OPEN ACCESS

Edited by:
Lucas Lima Verardo, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Brazil

Reviewed by:
Takashi Kohda, University of Yamanashi, Japan
Jianghan Qu, Veracyte, United States

***Correspondence:**
Kichoon Lee
lee.2626@osu.edu

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

Received: *14 April 2022*
Accepted: *21 June 2022*
Published: *22 July 2022*

Citation:
Ahn J, Lee J, Kim D-H, Hwang I-S, Park M-R, Cho I-C, Hwang S and Lee K (2022) Loss of Monoallelic Expression of IGF2 in the Adult Liver Via Alternative Promoter Usage and Chromatin Reorganization. Front. Genet. 13:920641. doi: 10.3389/fgene.2022.920641

In mammals, genomic imprinting operates *via* gene silencing mechanisms. Although conservation of the imprinting mechanism at the *H19/IGF2* locus has been generally described in pigs, tissue-specific imprinting at the transcript level, monoallelic-to-biallelic conversion, and spatio-temporal chromatin reorganization remain largely uninvestigated. Here, we delineate spatially regulated imprinting of *IGF2* transcripts, age-dependent hepatic mono- to biallelic conversion, and reorganization of topologically associating domains at the porcine *H19/IGF2* locus for better translation to human and animal research. Whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) of normal and parthenogenetic porcine embryos revealed the paternally hypermethylated *H19* differentially methylated region and paternal expression of *IGF2*. Using a polymorphism-based approach and omics datasets from chromatin immunoprecipitation sequencing (ChIP–seq), whole-genome sequencing (WGS), RNA-seq, and Hi-C, regulation of *IGF2* during development was analyzed. Regulatory elements in the liver were distinguished from those in the muscle where the porcine *IGF2* transcript was monoallelically expressed. The *IGF2* transcript from the liver was biallelically expressed at later developmental stages in both pigs and humans. Chromatin interaction was less frequent in the adult liver compared to the fetal liver and skeletal muscle. The duration of genomic imprinting effects within the *H19/IGF2* locus might be reduced in the liver with biallelic conversion through alternative promoter usage and chromatin remodeling. Our integrative omics analyses of genome, epigenome, and transcriptome provided a comprehensive view of imprinting status at the *H19/IGF2* cluster.

Keywords: imprinting, IGF2, pigs, alternative promoter usage, biallelic conversion, chromatin reorganization

1 INTRODUCTION

Genomic imprinting operates in mammals as an epigenetic mechanism that leads to parent-of-origin-specific monoallelic expression of a subset of genes, mostly in a cluster, *via* silencing of either parental allele (Reik and Walter, 2001; Ferguson-Smith, 2011; Barlow and Bartolomei, 2014). Paternal or maternal expression of imprinted alleles is essential for embryonic development, animal growth and behavior, and diseases related to abnormal loss of imprinting (LOI) (Peters, 2014; Tian, 2014; Tucci et al., 2019). The allele-specific silencing is either direct (e.g., DNA hypermethylation on promoters (Ahn et al., 2021a) or indirect (e.g., by antisense non-coding RNAs (Latos et al., 2012) and chromatin insulators (Bell and Felsenfeld, 2000; Hark et al., 2000) and become complex when multiple types of silencing simultaneously occur on transcript isoforms generated by alternative promoter usage (Hayward et al., 1998a; Hayward et al., 1998b; Peters et al., 1999; Plagge et al., 2004; Ahn et al., 2020c). Regarding the insulators, studies have extensively investigated the *H19/Igf2* locus and established the insulator-mediated organization that coordinately regulates them *via* the imprinting control region (ICR) (Bartolomei et al., 1991; Dechiara et al., 1991; Ferguson-Smith et al., 1991; Bartolomei et al., 1993; Thorvaldsen et al., 1998; Bell and Felsenfeld, 2000; Hark et al., 2000). This ICR upstream (5′) of *H19* is methylated only in the paternal allele (i.e., paternally imprinted), so that the insulator CCCTC-binding factor (CTCF) is prevented from binding to the paternal allele of ICR and subsequently, the enhancer downstream (3′) of *H19* communicates with the far upstream promoter of *Igf2* to drive paternal *Igf2* expression. Insulin-like growth factor 2 (*IGF2*) is a growth factor that plays a central role in fetal and postnatal growth. Transgenic overexpression of the paternally expressed *Igf2* gene increased fetal growth (Sun et al., 1997), and upregulation or downregulation of *IGF2 via* aberrant imprinting is associated with the overgrowth disorder Beckwith-Wiedemann syndrome and the growth retardation disorder Silver-Russell syndrome, respectively (Jacob et al., 2013). The maternally expressed *H19* gene is a negative regulator of growth and encodes a tumor suppressor (Hao et al., 1993; Yoshimizu et al., 2008). This counteraction between paternally and maternally expressed genes posited in the parental conflict theory regulates balanced growth (Moore and Haig, 1991; Haig, 2004). Although the insulator model for the *H19/IGF2* locus has been established in mice and humans (Barlow and Bartolomei, 2014; Nordin et al., 2014), tissue- and transcript-specific imprinting and changes in chromatin organization during development remain to be identified in detail. These identifications can be achieved in a comparative manner in mammals that serve as biomedical models and are agriculturally important livestock. Pigs are relevant models for translational research due to their anatomical, physiological, as well as genomic similarities with humans (Lunney et al., 2021). Although studies have described DNA methylation and gene expression regarding *IGF2* and *H19* in pigs and a general conservation in the imprinting mechanism (Li et al., 2008; Park et al., 2009; Braunschweig et al., 2011; Criado-Mesas et al., 2019), tissue-specific imprinting at the transcript level, age-dependent hepatic monoallelic-to-biallelic conversion (loss of monoallelic expression), and spatio-temporal chromatin reorganization remain largely uninvestigated.

The multi-layered epigenetic regulatory machineries responsible for DNA methylation, chromatin accessibility, histone modifications, and gene expression can be investigated using integrative omics approaches. As the gold standard for DNA methylation analysis, whole-genome bisulfite sequencing (WGBS) of animal models including parthenogenetic embryos has been utilized to identify differentially methylated regions (DMRs) (Clark et al., 2006; Ahn et al., 2020a; Ahn et al., 2020c; Ahn et al., 2021b; a;Morrison et al., 2021). To assess chromatin accessibility (Thurman et al., 2012) and capture open chromatin sites, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) has been used (Buenrostro et al., 2013). Transcriptionally active promoters are marked by histone H3 trimethylated at lysine 4 (H3K4me3) (Howe et al., 2017). A subset of genes has an extended H3K4me3 signal, which covers the gene body and overlaps with the active enhancer mark—histone H3 acetylated at lysine 27 (H3K27ac), consisting of the broad epigenetic domain (Beacon et al., 2021). Cell type and tissue-specific enhancers in non-coding regulatory regions serve as key cis-regulatory elements for gene expression (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Andersson et al., 2014; Coppola et al., 2016). During development, epigenetic modifications alter enhancer activities, as shown in H3K27ac enrichment followed by up-regulation of extracellular matrix genes which might reduce myogenic potential in aged skeletal muscle (Zhou et al., 2019). The formation of open chromatin regions and maintenance of enhancer elements are related to activation of tissue-specific genes (Xu et al., 2007; Wiench et al., 2011). Collectively, epigenetic modulations change chromatin structure and thereby alter DNA accessibility, which affects availability of enhancers and promoters to the transcriptional machinery. These epigenetic modifications might affect tissue-specific monoallelic gene expression within the *H19/IGF2* locus, which can be identified by profiling informative single nucleotide polymorphisms (SNPs) in genomic DNA and mRNA of the same individual (Ahn et al., 2021b; a). In addition, chromosomal conformation capture-based methods such as Hi-C have enabled unbiased identification of chromatin interactions across the genome (Lieberman-Aiden et al., 2009). The genome is partitioned into functional domains of different scale including megabase-long and evolutionarily conserved topologically associating domains (TADs) in which intra-domain chromatin interactions are frequent and cis-regulatory elements are coordinately regulated (Shen et al., 2012). The insulation score for genomic intervals along the chromosome is used to detect minima/valleys of insulation profile for areas of reduced chromatin interactions which are classified as TAD boundaries (Crane et al., 2015). As such, investigating the chromatin structure of imprinted domains in terms of TAD organizations (Lleres et al., 2019; Li et al., 2020) improves our understanding on imprinting clusters in the chromosomal context.

Here we aimed to delineate tissue-specific imprinting of major *IGF2* transcripts and hepatic mono- to biallelic conversion during development of pigs in comparison with humans. We found that the monoallelic-to-biallelic switch through liver-specific alternative promoter usage might occur concomitantly with removal of TAD boundaries and lower chromatin interaction frequencies at the porcine *H19/IGF2* locus. Our findings provided a comprehensive view of coordinated action of regulatory elements and chromatin organization and better understanding of tissue-specific and developmentally regulated genomic imprinting at the *H19/IGF2* locus.

2 MATERIALS AND METHODS

2.1 Ethics Statement

Our experimental protocols for parthenogenetic studies in the pig were approved by the Institutional Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration (RDA) of Korea (approval number NIAS 2015-670). Access to the National Bioscience Database Center (NBDC) Human Database for de-identified data was controlled to observe the Ohio State Human Research Protection Program (HRPP) policies on human subjects (study number 2020E1322).

2.2 Collection of Parthenogenetic and Control Embryos

Procedures of production of parthenogenetic embryos following *in vitro* maturation (IVM) of pig oocytes have been described in our previous reports (Kwon et al., 2017; Ahn et al., 2020b). In detail, ovaries of Landrace x Yorkshire x Duroc (LYD) pigs were obtained from a local slaughterhouse, transferred to our lab, and maintained in a thermos at 30–35°C. Cumulus-oocyte complexes (COCs) were gathered and washed in Tyrode's lactate-Hepes medium containing 0.1% (w/v) polyvinyl alcohol. Before IVM, 50 COCs were washed three times in TCM-199 (GIBCO, Grand Island, NY, United States) [supplemented with 0.1% polyvinyl alcohol (w/v), 3.05 mM D-glucose, 0.91 mM sodium pyruvate, 0.57 mM cysteine, 0.5 μg/ml luteinizing hormone, 0.5 μg/ml follicle stimulating hormone, 10 ng/ml epidermal growth factor, 10% porcine follicular fluid (pFF), 75 μg/ml penicillin G, and 50 μg/ml streptomycin] and then placed in each well of five 4-well dishes (Nunc, Roskilde, Denmark) containing 500 μL of maturation medium and matured for 40–42 h at 38.5°C in an incubator containing 5% CO_2. After maturation, cumulus cells were removed and oocytes having the first polar body were selected and activated as follows: oocytes were placed in a fusion chamber with 250 μm diameter wire electrodes (BLS, Budapest, Hungary) covered with 0.3 M mannitol solution containing 0.1 mM $MgSO_4$, 1.0 mM $CaCl_2$, and 0.5 mM Hepes and two DC pulses (1 s interval) of 1.2 kV/cm were applied for 30 μs using an LF101 Electro Cell Fusion Generator (Nepa Gene Co., Ltd. Chiba, Japan). Then, after 2 h of stabilization period, parthenogenetic embryos were placed into oviducts of two LY (Landrace X Yorkshire) surrogate gilts aged 12 months at onset of estrus to develop the embryos. Parthenogenetic embryos were recovered at day 21 from the surrogate gilts before they underwent morphological changes around day 30–35 (Bischoff et al., 2009; Hwang et al., 2020). As a control, fertilized embryos were also recovered at day 21 from gilts, after two LY gilts were naturally mated with boars twice with a 6 h interval during the natural heat period at the onset of estrus and confirmed pregnant by ultrasound examination. For the recovery, gilts were euthanized, and their reproductive tracts were sectioned, and the placenta was isolated from the uterus. Embryos were separated from the surrounding placenta and the surface of embryos was dried on cleaning tissues. Morphologically intact embryos with comparable sizes (approx. 2 cm) were selected and stored in liquid N2 until further use.

2.3 Whole-Genome Bisulfite Sequencing

Genomic DNA was isolated from whole collected embryos (triplicates for both parthenogenetic and control embryos) and fragmented. Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences, Inc. Ann Arbor, MI, United States) was used to optimize bisulfite conversion of genomic DNA according to the manufacturer`s instructions. PCR was conducted to amplify the bisulfite-treated DNA with adapter primers, Diastar™ EF-Taq DNA polymerase (Solgent, Daejeon, Korea), and the following thermal cycles: 3 m at 95°C followed by 35 cycles of 30 s at 95°C, 30 s at 60°C, and 30 s at 72°C, and a final extension for 5 m at 72°C. After bead-based clean-up, libraries were sequenced on an HiSeqX sequencer by Macrogen (Seoul, Korea) with 151 bp paired-end reads. Data quality was checked using FastQC (v0.11.7). Raw reads in FASTQ format were quality- and adapter-trimmed with the default parameters of Trim Galore (v0.4.5), except for additional trimming of 18 bp off the 3′ end of R1 and the 5′ end of R2 for removing bases derived from the sequence tag introduced in the library preparation procedure (--three_prime_clip_R1 18 --clip_R2 18). Trimmed reads (more than 800 million reads for each sample) were aligned to the pig reference genome (Sscrofa11.1, GenBank accession: GCF_000003025.6) using Bismark (v0.22.3) with default parameters including --no_overlap for paired-end reads (Krueger and Andrews, 2011). After deduplication using the deduplicate_bismark command, the Bismark methylation extractor was used to calculate methylation percentage of every cytosine in CpG context. Next, the DMR caller, metilene (v0.2-8), was used to identify *de novo* DMRs with default parameters [including maximum distance of 300 bp between CpGs (-M 300), minimum of 10 CpGs (-m 10), and minimum mean methylation difference of 0.2 (−d 0.2)] and a false discovery rate (FDR) option (Juhling et al., 2016). Methylation ratios and DMRs (FDR < 0.05) were visualized on genomic coordinates using the R/Bioconductor package Gviz (v1.28.3) (Hahne and Ivanek, 2016).

2.4 RNA Sequencing

Total RNA from whole embryo samples (n = 3 for each of the control and parthenote) was isolated with TRIzol reagent (Sigma-Aldrich, United States) following the manufacturer's instructions. The RNA samples were treated with DNase I to avoid genomic

DNA contamination and electrophoresed in 1.2% agarose gels to evaluate the integrity of RNA, which was confirmed by 28S/18S rRNA ratio > 2 and RNA integrity number (RIN) > 7 using an Agilent 2100 BioAnalyzer. The concentrations of RNA were assessed by the ratios of A260/A280 and A260/A230 (1.8–2.0). One ug of total RNA and the TruSeq RNA Sample Prep Kit v.2 (Illumina, San Diego, CA, United States) were used to construct cDNA libraries, and final libraries were produced using the protocol consisting of polyA-selected RNA extraction, RNA fragmentation, random hexamer primed reverse transcription and amplification. The cDNA libraries were quantified by quantitative Real-Time PCR (qPCR), and qualification of the libraries was assessed using the Agilent 2100 Bioanalyzer. The library products (100 nt paired-end) were sequenced by the Illumina HiSeq2500 RNA-Seq platform. The raw RNA sequencing reads (FASTQ format) were checked for quality by FastQC (v0.11.7) and trimmed and filtered by Trimmomatic v0.38 with default parameters (Bolger et al., 2014). Then, using STAR aligner (v2.7.5) with default parameter settings (Dobin et al., 2013), cleaned sequencing reads were mapped to the pig reference genome sequence (Sscrofa11.1). Duplicated reads were removed using Picard MarkDuplicates and reads were filtered using SAMtools (-q 30) (Li et al., 2009). Read coverages in BAM files were normalized to values equivalent to transcripts per million (TPM) using bamCoverage in deepTools (v3.5.0) with parameters (--binSize 10, --smoothLength 15) (Ramirez et al., 2014) and plotted using the R/Bioconductor package Gviz (v1.28.3) (Hahne and Ivanek, 2016).

2.5 Analyses of Differential Gene Expression

Raw RNA-seq reads in the FASTQ format were quantified against indexed pig transcriptome using Salmon (v1.3.0) in the mapping-based mode (Patro et al., 2017). TPM values of each gene were obtained for parthenogenetic embryos (PA) and control embryos (CN) (n = 3 for each) from Salmon output files (quant.sf). The output files were then imported by tximport function to construct a gene-level DESeqDataSet object for the R/Bioconductor package DESeq2 (v1.28.1) (Love et al., 2014). The test for DEGs was conducted by DESeq2. To obtain significant DEGs, combined criteria of FDR < 0.05 and the absolute log2-fold change > 1 were used, where a fold change is defined as read counts in PA divided by read counts in CN.

2.6 Profiling Gene Regulatory Elements

Raw FASTQ files deposited with GEO accession number GSE158430 (ATAC-seq, and H3K27ac, H3K4me3, and CTCF ChIP-seq) for 6-month-old pigs (Kern et al., 2021), GSE143288 (ATAC-seq, and H3K27ac and H3K4me3 ChIP-seq) for 2-week-old pigs (Zhao et al., 2021), GSE153452 (CTCF-seq) for pig embryonic fibroblasts (Li et al., 2020), and GSE155324 (CTCF-seq) for human lymphoblasts (Ushiki et al., 2021) were downloaded *via* the European Nucleotide Archive (ENA) Globus GridFTP. The quality of the raw sequencing reads was checked using FastQC (v0.11.8), and raw reads were trimmed and filtered using Trimmomatic (v0.38) with default settings (Bolger et al., 2014). All trimmed reads were mapped to the pig reference genome (Sscrofa11.1) or the human reference genome (GRCh38.p13, RefSeq assembly accession: GCF_000001405.39) using BWA-MEM aligner (v 0.7.17-r1198) using default parameters (Li, 2013). For ATAC-seq, mitochondrial genome was removed from the pig genome before alignment to avoid contamination of the mitochondrial genome which is more accessible owing to chromatin packaging deficiencies (Yan et al., 2020). Aligned reads were deduplicated using Picard MarkDuplicates, and filtered for quality using SAMtools (MAPQ > 30) (Li et al., 2009). MACS2 was used with default parameters to call peaks except for broad peaks (--broad) for ATAC-seq and FDR < 0.01 (-q 0.01) (Zhang et al., 2008). Read coverages in BAM files were normalized to 1x depth (reads per genomic content, RPGC) using bamCoverage in deepTools (v3.5.0) with parameters (--binSize 10, --smoothLength 15) (Ramirez et al., 2014). Peaks were visualized on genomic coordinates using the R/Bioconductor package Gviz v1.36.2 (Hahne and Ivanek, 2016).

2.7 Analyses of Tissue-Specific and Developmental Stage-Specific Expression

Raw RNA-seq data of normal pigs [PRJEB44486 (under Sus scrofa section in the FAANG datasets, https://data.faang.org/dataset), GSE77776 (Li et al., 2017), PRJNA493166 (Zhang et al., 2019), GSE158430 (Kern et al., 2021), PRJNA597972, GSE124484, GSE92433, PRJNA721126, GSE93855 (Tang et al., 2017), and GSE157045 (Yang et al., 2021)] and humans [GSE63634 (Yan et al., 2016), SRP166862 (George et al., 2019), PRJNA395106, and GSE120795 (Suntsova et al., 2019)] were downloaded through ENA's Globus GridFTP, except for raw data files of the human adult liver (accession hum0158.v2 for controlled access) which were downloaded *via* SFTP of the NBDC Human Database. The RNA-seq processing procedures were the same as above.

2.8 Analyses of Allele-Specific Expression

Datasets with genomic DNA and mRNA sequencing data from the same individuals were used. In detail, in addition to the above RNA-seq data, raw data for genomic DNA from the same individual were also downloaded from ENA's Globus GridFTP, except for the human adult liver obtained from NBDC's SFTP. For each pig breed, one liver and skeletal muscle sample from a 60-day-old pig was used (PRJNA309108/GSE77776) (Li et al., 2017). The human fetal liver samples were from two fetuses at 12 weeks after gestation (GSE63634) (Yan et al., 2016). The IDs of human liver samples from adults aged 31–74 were RK001, RK003, RK018, RK019, RK024, RK075, RK130, RK141, and RK157 (hum0158.v2). The IDs of human smooth muscle samples from reproductive age adults were MP100N, MP136N, MP169N, NW206N, and GO537N (SRP163897/SRP166862) (George et al., 2019). The IDs of human lung samples of adults aged 68–77 were N1, N3, N5, N12, N19, and N23 (PRJNA395106). Individuals with these IDs had

heterozygous SNPs at the *IGF2* locus in genomic DNA. In addition, an RNA-seq dataset, PRJNA597972, was used to analyze biallelic tendencies.

Whole-genome sequencing (WGS), whole-exome sequencing (WES), and H3K4me1 ChIP-seq data were cleaned and aligned using BWA-MEM as above. The deduplicated BAM files were used to detect SNPs and obtain allele counts in individual samples by generating vcf files through bcftools mpileup piped to bcftools call command. The published pig SNP data (GCA_000003025.6_current _ids.vcf.gz) were obtained from the EBI ftp server (ftp://ftp.ebi.ac.uk/pub/databases/eva/rs_releases/release_2/by_species/sus_scrofa/Sscrofa11.1/). The human SNPs in the vcf file (GCF_000001405.39.gz) for the GRCh38.p13 genome were downloaded from the NCBI data repository (https://ftp.ncbi.nih.gov/snp/.redesign/latest_release/VCF/). For sequencing to generate chromatogram for *IGF2(8)* and *IGF2-AS*, genomic DNA and RNA were isolated from the liver of 6-month-old Berkshire pigs. To amplify a DNA fragment containing a potential SNP on the second exon of the *IGF2(8)* transcript, primers were designed on the first intron (forward: 5′-AGCGTGGA GAGGCTCTCTT-3′) and the second intron (reverse: 5′-ACCCAAACACTCAATGCAGCTTT-3′). For a potential SNP on the second exon of the *IGF2-AS* transcript, primers were designed on the first intron (forward: 5′-CTGCTC TGGGTTCCCCAT-3′) and the second intron (reverse: 5′-CTGACAA CCCTGCCCTGTT-3′). After sequencing and confirming heterozygosity of the SNP, primers for cDNA were designed on the first exon (forward: 5′-CCCCATTGGCACCAG TACAG-3′) and the third exon (reverse: 5′-GCTGAGCCCGAG GAGATGTG-3′) of *IGF2(8)* and the first exon (forward: 5′-GGA CACGCGAGGCGA-3′) and second exon (reverse: 5′-CAAGGT CCAGGCGCATGT-3′) of *IGF2-AS* to avoid genomic DNA contamination. PCR was conducted using the Taq DNA Polymerase (#M0273S, New England Biolabs, Ipswich, MA, United States) with an initial incubation at 95°C for 30 s, followed by 43 cycles at 95°C for 30 s, 56°C for 35 s, and 68°C for 20 s. The final extension was performed at 68°C for 5 min. After agarose gel electrophoresis, DNA was extracted from separated PCR bands using a QIAquick Gel Extraction Kit (#28104, Qiagen, Venlo, Netherlands) according to the manufacturer's protocol and sent out for Sanger sequencing at The Ohio State University Core Facility.

2.9 Hi-C Data Processing

Raw Hi-C data in FASTQ format of the liver of fetal (embryonic day 90) and adult (2-years-old) Bamaxiang pigs (PRJNA482496) (Tian et al., 2020) and skeletal muscle tissues of a Luchuan pig (embryonic day 35) (GSE166346) (Yuan et al., 2021) and Large White pigs (2-week-old) (GSE143288) (Zhao et al., 2021) were retrieved through ENA's Globus GridFTP. After assessing the quality of data using FastQC (v0.11.7), the raw paired-end reads were trimmed and filtered out to remove low quality reads, adapters, and reads shorter than 20 bp by using default settings of Trim Galore (v0.4.5). Cleaned data were processed using HiC-Pro (v.3.1.0) with default parameters (Servant et al., 2015) while specifying the index for bowtie2 (v2.4.4) and MboI (or DpnII) restriction fragments according to data submitters' publications (Tian et al., 2020; Yuan et al., 2021; Zhao et al., 2021). To determine concordance of raw Hi-C matrices, GenomeDISCO was used to produce smoothed matrices and randomly work on the smoothed matrices to obtain concordance scores (Ursu et al., 2018). The validPairs files from matrices with high concordance were merged to increase resolution and normalized by iterative correction and eigenvector decomposition (ICE) using parameters of HiC-Pro (-s merge_persample -s build_contact_maps -s ice_norm). TADs were identified using insulation scores and ICE normalized matrices were visualized using the GENOVA R-package (Van Der Weide et al., 2021).

3 RESULTS

3.1 A Differentially Methylated Region Within the Porcine *H19/IGF2* Locus is Paternally Methylated at a CpG Island

Using diploid uni-maternal PA embryos and bi-parental CN embryos, we performed WGBS at an approximately 50X depth (**Supplementary Table S1**). By analyzing the WGBS data, DMRs between the embryos were obtained whose mean methylation difference (i.e., a mean of PA subtracted by a mean of CN) was more than 0.2 (hypermethylation in PA) or less than 0.2 (hypomethylation in PA) with significance (FDR < 0.05) (**Supplementary Figure S1A**). Compared with methylated regions without significance, DMRs tended to be longer in base pairs and greater in CpG numbers, as a result of processing approximately 588 million deduplicated uniquely mapped reads for each replicate on average (**Supplementary Figure S1B** and **Supplementary Table S1**). Using unmethylated lambda phage DNA added to sample DNA prior to fragmentation, bisulfite conversion rates at CpG, CHG, and CHH sites were calculated and estimated to be 99.69%–99.71% across samples representing successful library construction (**Supplementary Table S1**). In order to examine DNA methylation status within the *H19*/*IGF2* locus in porcine embryos, methylation ratios at single-base resolution were analyzed. Based on the mean methylation ratio at each CpG site, diploid uni-maternal PA carried a broad range of hypomethylation immediately upstream (5′) of the non-coding *H19* gene (**Figure 1A** bottom panel, **Supplementary Table S2**). Hemi-methylation occurred in bi-parental CN, and a DMR between PA and CN was identified which mostly overlapped a CpG island (**Figure 1A** bottom panel). It was consistent with a previous study regarding the presence of the porcine *H19*-DMR, where maternal alleles were almost unmethylated and paternal alleles were completely methylated which led to hemi-methylation (Braunschweig et al., 2011). Additionally, a narrow hypermethylated DMR in PA was located farther upstream of *H19* (**Figure 1A** bottom panel). DNA methylation in the upstream and downstream of the *IGF2*, *INS*, *TH* genes did not show differences between PA and CN, and various *IGF2* transcripts did not exhibit a transcript-specific methylation pattern (**Figure 1A** bottom panel). In summary, the hypomethylated region in PA (i.e., paternally methylated *H19*-

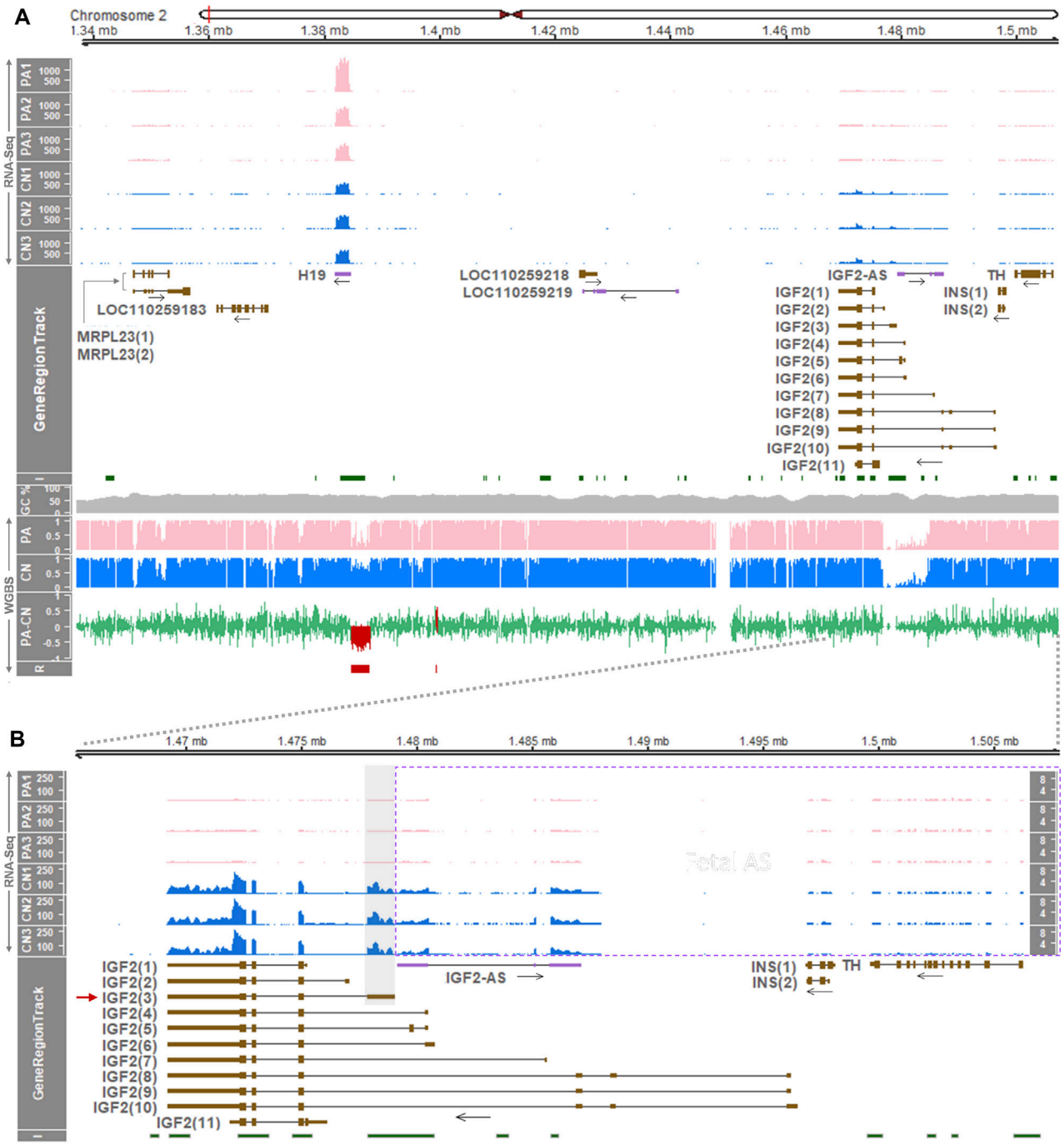


FIGURE 1 | Gene expression and DNA methylation profiling at the *H19/IGF2* locus in the porcine embryo. **(A)** Above GeneRegionTrack, normalized read coverages from RNA-Seq are shown as transcripts per million (TPM; y-axis) in each track of PA and CN embryos. In GeneRegionTrack, genes located in the 0.171-Mb (171-Kb) region between the *MRPL23* and *TH* genes (chr2:1,337,000-1,508,000) are displayed by either brown boxes [tall, translated regions; short, untranslated regions (UTRs)] for protein-coding transcripts or purple boxes for non-coding transcripts. The directions of transcription are marked by horizontal arrows. Below GeneRegionTrack, averages of cytosine methylation ratios (n = 3, PA and CN each) obtained by WGBS are shown. PA-CN indicates mean methylation ratios of PA subtracted by those of CN. R represents DMRs called by the metilene software (FDR < 0.05) which are also overlaid in the PA-CN track. I (CpG islands) and GC% (GC content) were derived from the UCSC Table Browser. **(B)** The locus spanning *IGF2*, *IGF2-AS*, *INS*, and *TH* genes (43-Kb, chr2:1,465,000-1,508,000) were zoomed in. For *IGF2-AS*, *INS*, and *TH*, expression is displayed within a dotted rectangle and an additional y-axis (TPM) is shown at the far right. Pig gene transcripts include *MRPL23(1)*, XM_021083608.1; *MRPL23(2)*, XM_021083607.1; *IGF2(1)*, XM_021080593.1; *IGF2(2)*, XM_021080637.1; *IGF2(3)*, XM_021080576.1; *IGF2(4)*, XM_021080612.1; *IGF2(5)*, XM_021080582.1; *IGF2(6)*, XM_021080648.1; *IGF2(7)*, XM_021080641.1; *IGF2(8)*, XM_021080607.1; *IGF2(9)*, XM_021080643.1; *IGF2(10)*, XM_021080603.1; *IGF2(11)*, XM_021080596.1; *INS(1)*, XM_021081278.1; *INS(2)*, NM_001109772.1.

DMR) was the only DMR that overlapped a CpG island throughout the *H19/IGF2* locus of the pig embryo, and transcript-specific DMRs were not found in the *IGF2* locus.

3.2 Expression of the *IGF2* Gene in Pig Embryos is Paternal Allele-Specific

In our model of PA and CN, gene and transcript expressions within the *H19/IGF2* locus were examined to investigate their imprinting status. Analyses of differentially expressed genes (DEGs) on RNA-seq data (**Supplementary Table S1**) revealed that expression of the non-coding *H19* gene tended to increase in PA compared to CN (1.61-fold higher in PA) suggesting its maternal expression, i.e., had a higher expression in two maternal alleles of PA than in one maternal allele of CN (**Figure 1A** top panel, **Supplementary Table S3**, and **Supplementary Figure S1C**). This deviation from a 2-fold increase might be accounted for by gene dosage compensation in diploid uni-parental embryos or loss of imprinting (Shemer et al., 1996; Park et al., 2011). Expression of the *IGF2* genes was almost exclusive in CN embryos (adjusted p-value < 0.001), indicating expression in the paternal allele of CN while being absent in PA without the paternal allele (**Figure 1B**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Among the *IGF2* transcript isoforms, the major transcript was *IGF2(3)* (short-form; GenBank accession number: XM_021080576.1) having four exons and its non-overlapping first exon carried predominant read coverages (**Figure 1B**). In addition, paternal expression of antisense of *IGF2* (*IGF2-AS*) was indicated by almost exclusive read coverages in CN (adjusted p-value < 0.001) (**Figure 1B**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Other genes and gene transcripts including *MRPL23* and *INS* appeared to be expressed biallelically and expression of the *TH* gene tended to increase in PA embryos, while expression of LOC110259183, LOC110259218, and LOC110259219 was almost undetectable (**Figure 1**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Consequently, the imprinted expressions of the *IGF2(3)* and *IGF2-AS* genes were shown to be paternal monoallelic.

3.3 Regions Surrounding the Porcine *H19* and *IGF2* Locus Accumulate Distinguishable Gene Regulatory Elements Between the Liver and Skeletal Muscle

To examine whether imprinted monoallelic expression at the *IGF2* locus is maintained in different developmental stages, we first analyzed expression levels *H19* and *IGF2* in multiple tissues of 6-month-old pigs using a dataset retrieved by the Gene Expression Omnibus (GEO) accession GSE158430. At this stage, expression of *IGF2* was prominent in the liver and expression of *H19* was high in skeletal muscle (**Supplementary Figure S2**). In other pig tissues, including the adipose tissue, brain hypothalamus, lung, and spleen, expression of both *IGF2* and *H19* was substantially low (**Supplementary Figure S2**). Therefore, we investigated gene regulatory elements in the liver and skeletal muscle of the same pigs within and near the porcine *H19/IGF2* locus that might affect gene expression. ATAC-seq (for open chromatin), H3K27ac (at active enhancers and promoters), H3K4me3 (at active promoters), and CTCF (insulators) were analyzed. Our analysis of these datasets from two biological replicates of skeletal muscle revealed that H3K27ac signals were distributed around ATAC peaks spanning ~20 kb (chr2:1,346,472-1,366,301) approximately 15–35 kb downstream (3′) of the *H19* transcription start site (TSS) (ae, **Figure 2A**), suggestive of active enhancers. However, these signals were absent in the liver. On the other hand, H3K27ac peaks upstream (5′) of *IGF2(8)* were enriched with H3K4me3 in the liver, and also H3K27ac peaks upstream of *IGF2(3)* were enriched with H3K4me3 in skeletal muscle (approx. 1–2 kb H3K27ac and H3Kme3 signals) (**Figure 2A**), suggesting that they represent regulatory features including active promoters. In addition, in skeletal muscle but not in the liver, H3K4me3 signals located immediate upstream (5′) of *H19* where CTCF peaks co-localized (ap, **Figure 2A**). It indicated that the *H19* promoter was activated, and this activation might be attributed to the aforementioned active enhancer downstream of *H19* and monoallelic CTCF binding which leads to monoallelic activation of the *H19* promoter as previously reported regarding the human and mice ICRs (Bell and Felsenfeld, 2000; Hark et al., 2000). Moreover, although monoallelic CTCF binding in the skeletal muscle was not confirmed due to lack of heterozygous sites, the monoallelic CTCF binding was evident in pig embryonic fibroblasts (PEFs) whose CTCF sites were analogous to those of human lymphoblasts (**Supplementary Figure S3**). Noticeably, expression of *H19* was low in the liver and high in skeletal muscle (**Figure 2A** bottom panel, and **Supplementary Figure S2**), and *IGF2* transcripts expressed in this locus were different between the two tissues: *IGF2(8)* in the liver and *IGF2(3)* in skeletal muscle, and total expression of *IGF2* was higher in the liver (**Figure 2A** bottom panel, **Figure 2B**, and **Supplementary Figure S2**).

We further investigated an additional presence of gene regulatory elements in the liver which deviate from the insulator model (Barlow and Bartolomei, 2014; Nordin et al., 2014) using another dataset (GSE143288). In 2-week-old pigs, H3K27ac peaks were distributed in the downstream of *H19* around the open chromatin region indicated by ATAC signals, although the dataset contained ATAC data for the one pig (Li1) (**Figure 3A** top panel). The H3K4me3 signals near *H19* indicated an activated promoter, and all the ATAC, H3K27ac, and H3K4me3 signals around the first exon of *IGF2(3)* suggested the long-range insulator-mediated regulation of *IGF2(3)* expression. However, those ATAC, H3K27ac, and H3K4me3 signals were also detected near the first exon of the *IGF2(8)* transcript (**Figure 3A** top panel), suggesting the presence of an additional transcript-level gene regulation. Consistently, the expression of both *IGF2(3)* and *IGF2(8)* was observed in the same liver tissues while *H19* was expressed in the far downstream of *IGF2* (**Figure 3A** bottom panel and **Figure 3B**). Taken together, it suggested that, compared to the *IGF2(3)* expression in the whole embryo (**Figure 1**) and the *IGF2(8)* expression in the liver from 6-month-old pigs (**Figure 2**), both *IGF2(3)* and *IGF2(8)* are

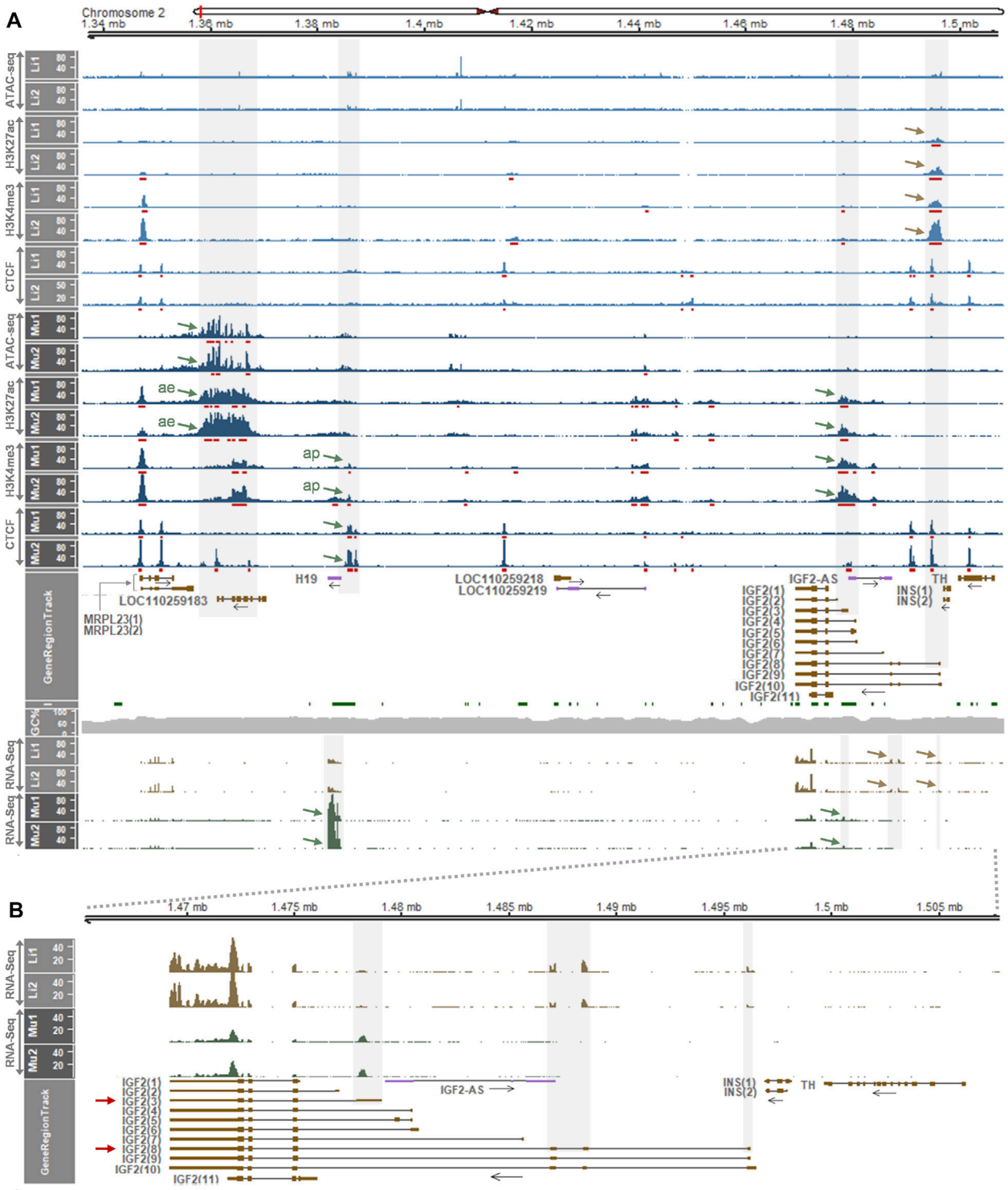


FIGURE 2 | Epigenetic regulatory elements identified in the *H19/IGF2* locus of liver and skeletal muscle of 6-month-old pigs and gene expression profiling. **(A)** Above GeneRegionTrack, peaks of ATAC-seq, H3K27ac, H3K4me3 and CTCF are displayed in 1x normalized read coverages. MACS2-called peaks are underscored with red bars. Peaks of interest in the liver and skeletal muscle are pointed with brown and blue arrows, respectively. ae, active enhancer; ap, active promoter. Below GeneRegionTrack, normalized read coverages in TPM values from RNA-Seq from the same pigs (P348 and P350) are displayed. Li1, liver 1 from P348; Li2, liver 2 from P350; Mu1, skeletal muscle 1 from P348; Mu2, skeletal muscle 2 from P350. **(B)** RNA-Seq read coverages in the *IGF2* locus are zoomed in. Data were retrieved from a dataset (GSE158430). Details about y axis titles and the plot are in **Figure 1** legend.

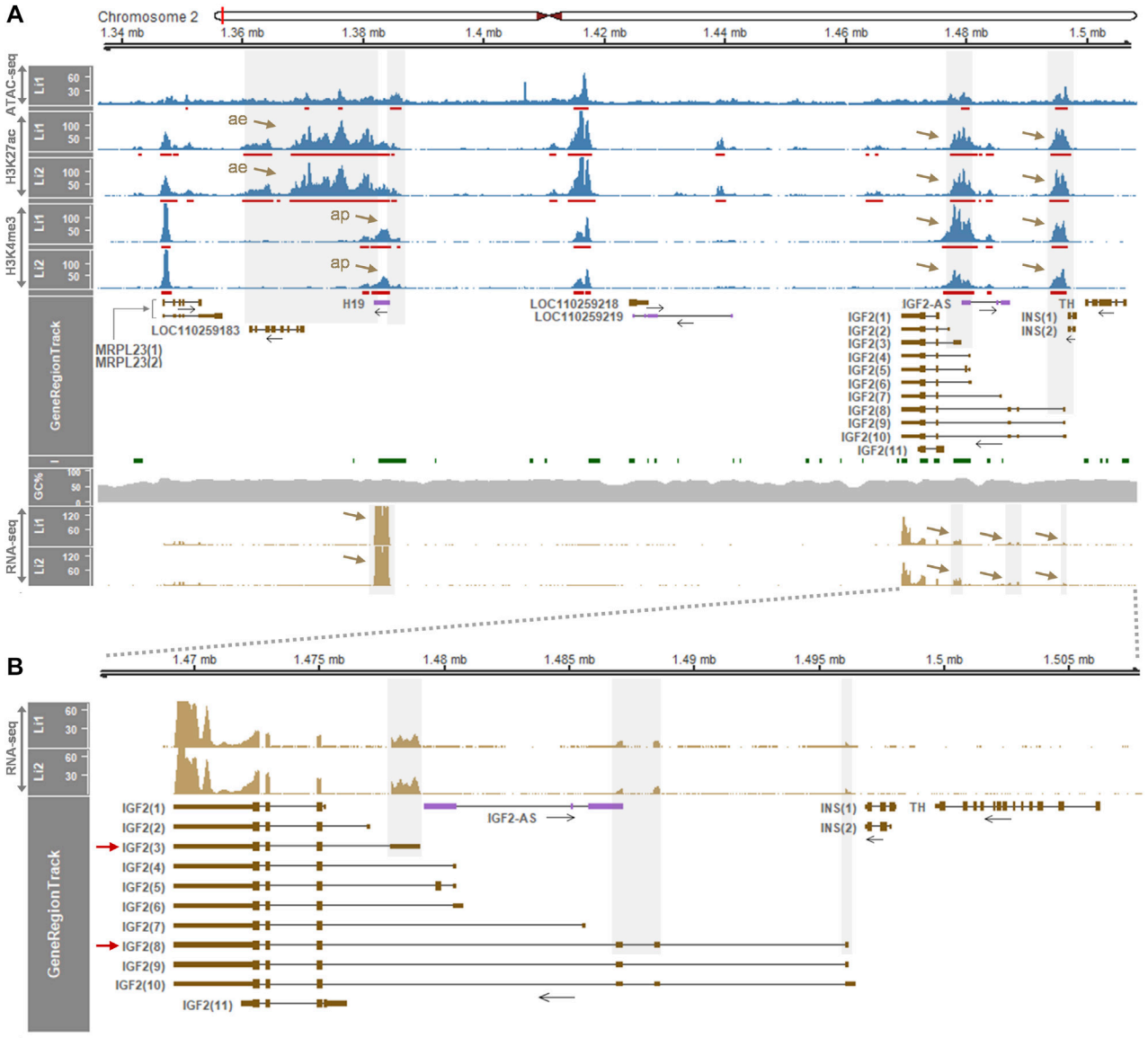


FIGURE 3 | Enrichment of gene regulatory signals and expression of genes within the *H19/IGF2* locus of the liver of 2-week-old pigs. **(A)** ATAC-seq, H3K27ac, and H3K4me3 signals in the liver (Li) of two Large White pigs are represented with 1x normalized read coverages. Red bars indicate MACS2-called peaks. Both grey shades and brown arrows denote gene regulatory signals and gene expression. ae, active enhancer; ap, active promoter. **(B)** Expression pattens of the *IGF2* transcripts is zoomed in. Both *IGF2(3)* and *IGF2(8)* are expressed in the same liver tissues of the 2-week-old pigs. Data were retrieved from GSE143288.

expressed in the liver of two-week-old pigs (**Figure 3**) under the regulation of two different sets of gene regulatory elements.

3.4 Developmental Changes of the Porcine *IGF2* Gene Expression and Bi- or Mono-Allelic Expression are Distinct in the Liver and Skeletal Muscle

Considering the aforementioned distinct gene regulation and expression patterns in the liver, we examined whether the *IGF2* gene expression is regulated developmentally and shows an allele-specific pattern. First, using RNA-seq datasets, read coverages for *IGF2* transcripts were analyzed in the liver and skeletal muscle across developmental stages. In both the liver and skeletal muscle, on embryonic day 70 and postnatal day 1, the *IGF2(3)* transcript was predominantly expressed (**Figure 4**). On day 60, however, the *IGF2(3)* transcript was not always predominant in the liver of analyzed pigs, while in the skeletal muscle the *IGF2(3)* transcript was predominant (**Figure 4**). In particular, in the liver of Landrace pigs, the *IGF2(3)* transcript was predominant and the *IGF2(8)* expression was detected at a low level. In the liver of Large White and Meishan pigs both *IGF2(3)* and *IGF2(8)* transcripts were predominant, while Berkshire pigs showed high expression of the *IGF2(8)* transcript and reduced

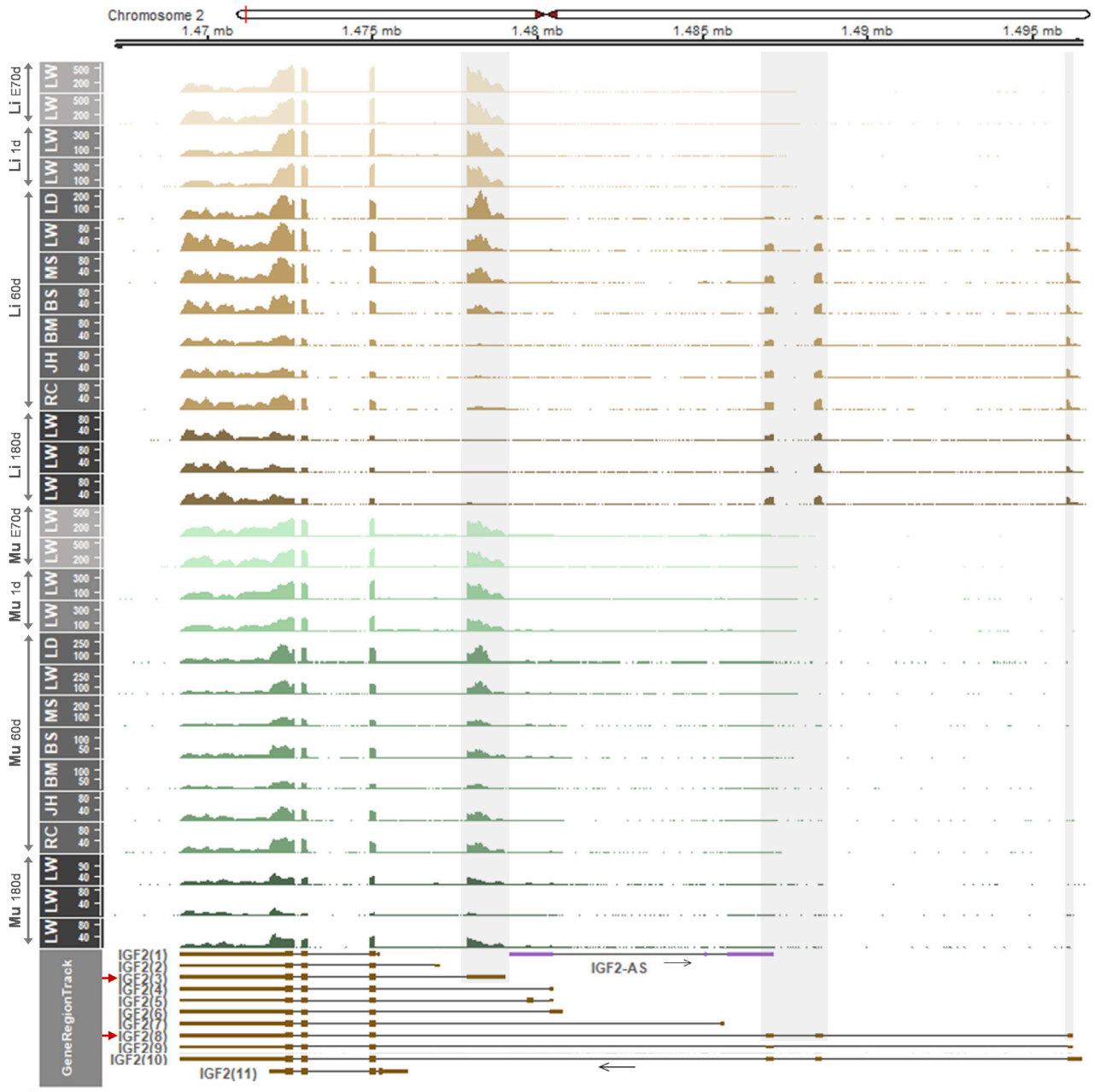


FIGURE 4 | mRNA expression in the liver and skeletal muscle within the *IGF2* locus of embryonic day 70 (E70 d), 1-day-old (1 d), 60-day-old (60 d), and 180-day-old (180 d) pigs. Pig RNA-seq data generated using embryonic liver and skeletal muscle (PRJEB44486), 1 d liver and skeletal muscle (PRJEB44486), 60 d liver and skeletal muscle (GSE77776), and 180 d liver and skeletal muscle (PRJNA493166) were analyzed. Developmental stages are separated and indicated for liver and skeletal muscle using light brown to dark brown and light green to dark green, respectively. Non-overlapping exons with high read coverages are overlaid with grey shades and corresponding predominantly expressed transcripts (*IGF2(3)* and *IGF2(8)*) are marked by red arrows. Li, liver; Mu, skeletal muscle. LW, Large White; LD, Landrace; MS, Meishan; BS, Berkshire; BM, Bamei; JH, Jinhua; RC, Rongchang pigs.

expression of the *IGF2(3)* transcript. On the other hand, expression of the *IGF2(3)* transcript was almost absent in the liver of Bamei, Jinhua, and Rongchang pigs, whereas the *IGF2(8)* transcript was predominant (**Figure 4**). On day 180, the *IGF2(8)* transcript was predominant in the liver and the *IGF2(3)* transcript was predominant in skeletal muscle (**Figure 4**). The total read coverages of *IGF2* tended to decrease in both the liver and skeletal muscle during development. The transition of *IGF2* expression in the liver, but not in skeletal muscle, was repeatedly shown in additional data (**Supplementary Figures S4A and S5**).

To examine allelic expression patterns, informative (heterozygous) SNPs in the *IGF2(3)* and *IGF2(8)* transcripts

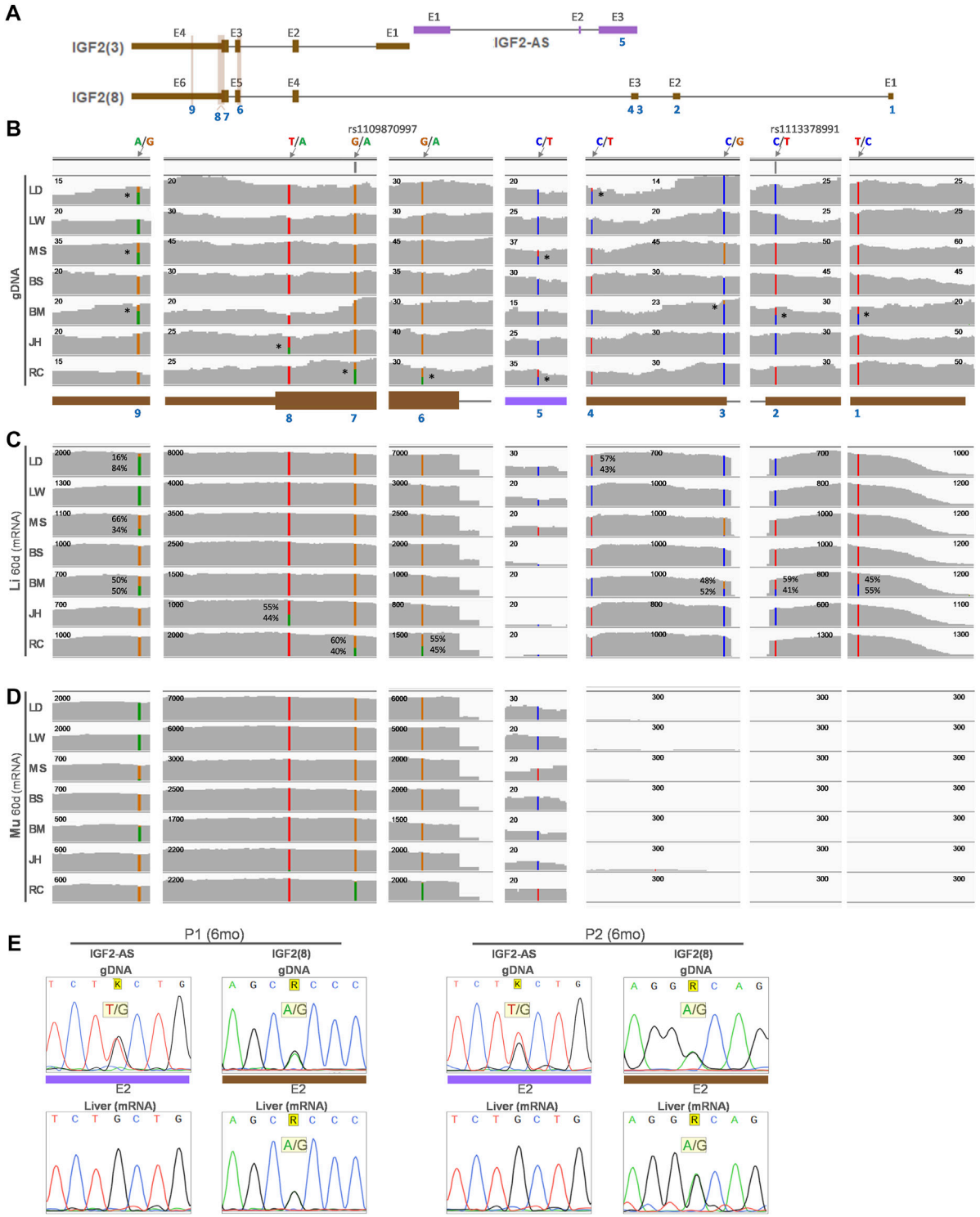


FIGURE 5 | Allelic expression of the *IGF2* transcripts from the liver and skeletal muscle of 60-day-old pigs from 9 breeds. **(A)** Predominant porcine *IGF2* transcripts in liver (Li) and skeletal muscle (Mu) from **Figure 4** are displayed along with an antisense transcript (*IGF2-AS*). **(B)** Location of SNPs on genomic DNA (gDNA) analyzed using WGS (PRJNA309108) are denoted with blue numbers. Reference and alternative alleles are marked in the format of ref/alt (e.g., A/G). If present, the reference SNP ID (rs ID) is shown (e.g., rs1109870997). Heterozygous alleles are denoted with stars (*) on the right side of SNPs in the genomic DNA. Read coverages of RNA-seq (GSE77776) from the same pigs as WGS are displayed for liver **(C)** and skeletal muscle **(D)**. Numbers on the top-left corner of each read coverage denote the depth of coverage. **(E)** Sequencing of exon 2 (E2) of *IGF2(8)* and *IGF2-AS* for genomic DNA and cDNA from the liver tissues of two 6-month-old (6 months) Berkshire pigs (P1 and 2). LW, Large White; LD, Landrace; MS, Meishan; BS, Berkshire; BM, Bamei; JH, Jinhua; RC, Rongchang pigs.

and *IGF2-AS* were explored (**Figures 5A,B**). Nine heterozygous SNPs were found in genomic DNA (gDNA) of any of the 60-day-old pigs (**Figures 5A,B** and **Supplementary Table S4**). Among those SNPs, four heterozygous SNPs including a previously reported SNP (rs1113378991) were found for non-overlapping exons of the *IGF2(8)* transcript in Bamei (SNP1-3) and Landrace (SNP4) pigs (**Figure 5B** and **Supplementary Table S4**) and mRNA expression on those four alleles in the liver was biallelic (**Figure 5C** and **Supplementary Table S4**), indicating biallelic expression of the porcine *IGF2(8)* transcript. One heterozygous SNP for *IGF2-AS* (SNP5) in Meishan and Rongchang pigs (**Figure 5B** and **Supplementary Table S4**) was expressed monoallelically in both the liver and skeletal muscle, although the mRNA expression level was low (**Figures 5C,D** and **Supplementary Table S4**). Four other SNPs (SNP6-9) including a previously reported SNP (rs1109870997) were found in overlapping exons of the *IGF2(3)* and *IGF2(8)* transcripts, and informative SNPs were not found in the non-overlapping exon (E1) of *IGF2(3)* (**Figures 5A,B** and **Supplementary Table S4**). In the liver of Rongchang, Jinhua, and Bamei pigs which expressed exclusively the *IGF2(8)* transcript (**Figure 4**), SNP6 and 7 in Rongchang, SNP8 in Jinhua, and SNP9 in Bamei pigs were biallelically expressed (**Figure 5C** and **Supplementary Table S4**). In the liver of Meishan pigs which expressed both the *IGF2(3)* and *IGF2(8)* transcripts (**Figure 4**), SNP9 showed a decreased biallelic tendency due possibly to monoallelic expression of *IGF2(3)* and biallelic expression of *IGF2(8)* (**Figure 5C** and **Supplementary Table S4**). In the liver of Landrace pigs which expressed the *IGF2(3)* transcript predominantly and the *IGF2(8)* transcript at a low level (**Figure 4**), SNP9 showed a monoallelic tendency due possibly to substantially higher monoallelic expression of the *IGF2(3)* transcript than biallelic expression of the *IGF2(8)* transcript (**Figure 5C** and **Supplementary Table S4**). These allelic changes were also found using another dataset which showed a biallelic tendency or a decreased biallelic tendency (**Supplementary Figure S4B**). In the skeletal muscle, these Rongchang, Jinhua, Bamei, Meishan, and Landrace pigs expressed the *IGF2(3)* transcript predominantly (**Figure 4**), and corresponding expression of SNP6-9 tended to be monoallelic (**Figure 5D** and **Supplementary Table S4**). In the liver of adult Berkshire pigs, biallelic expression of the *IGF2(8)* transcript and monoallelic expression of *IGF2-AS* were confirmed (**Figure 5E**). Taken together, it suggests that allelic expression patterns were different in between the liver and muscle, in addition to the difference of the expressed form of *IGF2* transcripts.

3.5 *IGF2* Expression in Humans is Developmentally Regulated and Bi- or Mono-Allelic Expression Patterns Are Tissue-Specific

For comparison with pigs, the *IGF2* gene expression in the fetal and adult liver and other tissues of the human was examined and also allelic expression in those tissues was further investigated. Based on RNA-seq read coverages, when compared with other tissues, expression of *IGF2* and *H19* was relatively high in the liver and muscle, respectively, while expression of both *IGF2* and *H19* was low in other tissues including the brain, lung, colon, and stomach (**Supplementary Figure S6A**). The *IGF2(2)* transcript [the orthologous transcript of porcine *IGF2(3)*] was expressed in the fetal liver and the *IGF2(6)* transcript [the orthologous transcript of porcine *IGF2(8)*] was expressed in the adult liver (**Figure 6** and **Supplementary Figures S6, S7**), which are comparable to the findings of the predominant expression of *IGF2(3)* and *IGF2(8)* in fetal and adult livers, respectively, in pigs. Both the *IGF2(1)* and *IGF2(2)* transcripts were expressed in the adult smooth muscle and lung (**Figure 6**). The overall expression levels tended to be high to low in the order of fetal liver, adult liver, adult smooth muscle, and adult lung. Expression of *INS-IGF2* fusion transcripts and *INS* transcripts was not detectable (**Figure 6**), while expression of *IGF2-AS* was low but detectable in the fetal liver and decreased in postnatal stages of both pigs and humans (**Figure 7B** and **Supplementary Figure S8**). Total expression of *IGF2* was significantly higher in the liver than in skeletal muscle, and total expression of *H19* tended to be higher in skeletal muscle than in the liver (**Supplementary Figure S6B**), and these patterns in adult humans were similar to those of the adult pigs (**Supplementary Figure S2B**).

Allelic expression was further explored for the fetal and adult liver of the human, regarding informative SNPs in the expressed *IGF2(2)* and *IGF2(6)* transcripts and *IGF2-AS* (**Figure 7A**). The *IGF2(2)* and *IGF2-AS* transcripts were expressed in the fetal liver (**Figures 7A,B**), and six informative SNPs on gDNA were found (**Figure 7B** and **Supplementary Table S5**). All of those six SNPs were monoallelically expressed in mRNA of the fetal liver (**Figure 7B** and **Supplementary Table S5**). The *IGF2(6)* transcript was expressed in the adult liver and with a very low degree for *IGF2-AS* expression (**Figures 7A,C**). The same SNP5 between the fetal and adult liver was heterozygous in the adult liver from individual 2, 4, 6, 7, and 9 (**Figure 7C** top panel and **Supplementary Table S5**) and tended to be expressed biallelically (**Figure 7C** bottom panel and **Supplementary Table S5**), suggesting biallelic expression of *IGF2(6)* unlikely to monoallelic expression of *IGF2(2)* in the fetal liver (**Figure 7B** and **Supplementary Table S5**). Also, in the adult liver, SNP8-10 were found to be heterozygous in some individuals (SNP8: 2 and 4-9, SNP9: 2, 6, 7, and 9, SNP10: 2, 4, 7, and 9) and corresponding mRNA expression of *IGF2(6)* tended to be biallelic (**Figure 7C** and **Supplementary Table S5**). The SNP7 was found on *IGF2-AS* but the expression was too low to determine its allelic expression (**Figures 7A,C** and **Supplementary Table S5**) and therefore, its allelic expression could not be determined.

Moreover, in smooth muscle and lungs, informative SNPs on *IGF2(1)*, *IGF2(2)*, and *IGF2-AS* were explored in gDNA (**Figure 8A**) and their allelic expression was examined. In smooth muscle, both *IGF2(1)* and *IGF2(2)* were expressed, and SNP11-15 were heterozygous in gDNA (**Figures 8A,B** and **Supplementary Table S6**). Allelic expression of SNP11-15 tended to be monoallelic (**Figure 8B** bottom panel and **Supplementary Table S6**), indicating monoallelic expression of *IGF2(1)* and *IGF2(2)*. The SNP1 was heterozygous in smooth

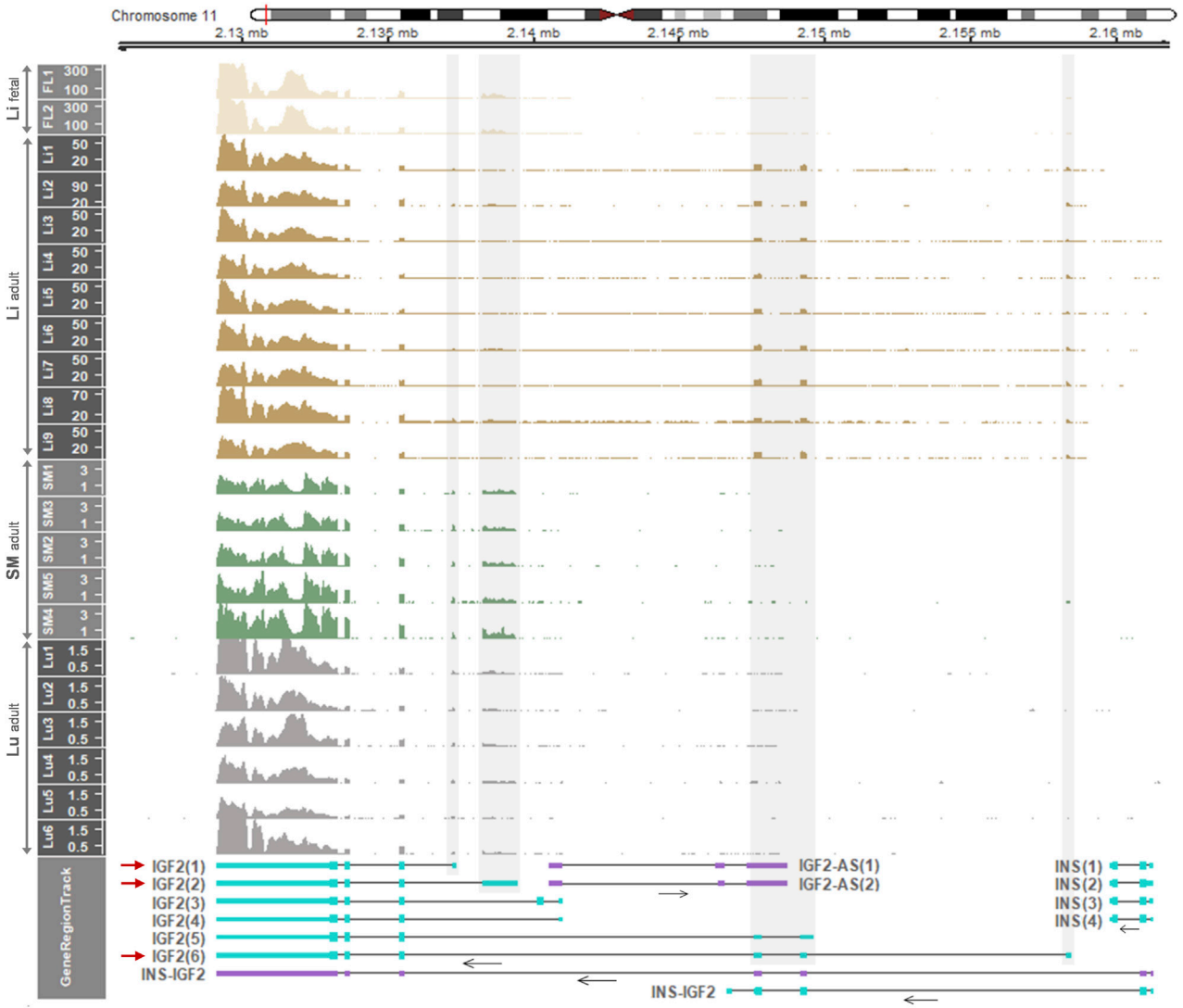


FIGURE 6 | Temporal and spatial differences of expressed *IGF2* transcripts in the human. Displayed are mRNA expression of the *IGF2* gene in the human fetal liver (Li), adult liver (Li), adult smooth muscle (SM), and adult lung (Lu). RNA-seq data were retrieved from datasets: GSE63634 for fetal liver, hum0158.v2 for normal adult liver, SRP166862 for normal adult smooth muscle, and PRJNA395106 for normal adult lung. Read coverages are represented as TPM values on the y-axis. In GeneRegionTrack, transcripts of *IGF2*, *INS-IGF2* fusion, *IGF2-AS*, and *INS* genes located in the 0.0365-Mb (36.5-Kb) region (chr11:2,125,500-2,162,000) are displayed by either light blue boxes [tall, translated regions; short, untranslated regions (UTRs)] for protein-coding or purple boxes for non-coding. Horizontal arrows under transcripts indicate the direction of transcription. Grey shades overlay non-overlapping exons with high read coverages corresponding to predominant transcripts (*IGF2(1)*, *IGF2(2)* and *IGF2(6)*) with red arrows. Human gene transcripts include *IGF2(1)*, NM_001291861.3; *IGF2(2)*, NM_000612.6; *IGF2(3)*, NM_001127598.3; *IGF2(4)*, NM_001291862.3; *IGF2(5)*, NM_001007139.5; *IGF2-AS(1)*, NR_028043.2; *IGF2-AS(2)*, NR_133657.1; *INS(1)*, NM_001185097.2; *INS(2)*, NM_001185098.2; *INS(3)*, NM_001291897.2; *INS(4)*, NM_000207.3.

muscle, but the mRNA expression was very low (**Figures 8A,B** and **Supplementary Table S6**). In the lungs, both *IGF2(1)* and *IGF2(2)* were expressed, and the same SNP5, 9, and 10 as in the liver were found to be heterozygous (**Figures 8A,C** and **Supplementary Table S6**). Unlike biallelic expression in the liver, expression on SNP5, 9, and 10 was monoallelic in the lungs (**Figure 8C** bottom panel and **Supplementary Table S6**). In addition, SNP16 was found to be heterozygous in individual 2, and its expression tended to be monoallelic in the lungs (**Figures 8A,C** and **Supplementary Table S6**). Regarding SNP1, the mRNA expression of *IGF2-AS* was very low in the lungs (**Figures 8A,C** and **Supplementary Table S6**).

Overall, in the human, *IGF2(1)* and/or *IGF2(2)* were expressed in the fetal liver, adult smooth muscle, and adult lung, and the expression tended to be monoallelic, whereas *IGF2(6)* was expressed in the adult liver and the expression tended to be biallelic. In addition, the *IGF2-AS* expression was very low, except in the fetal liver where the expression tended to be monoallelic.

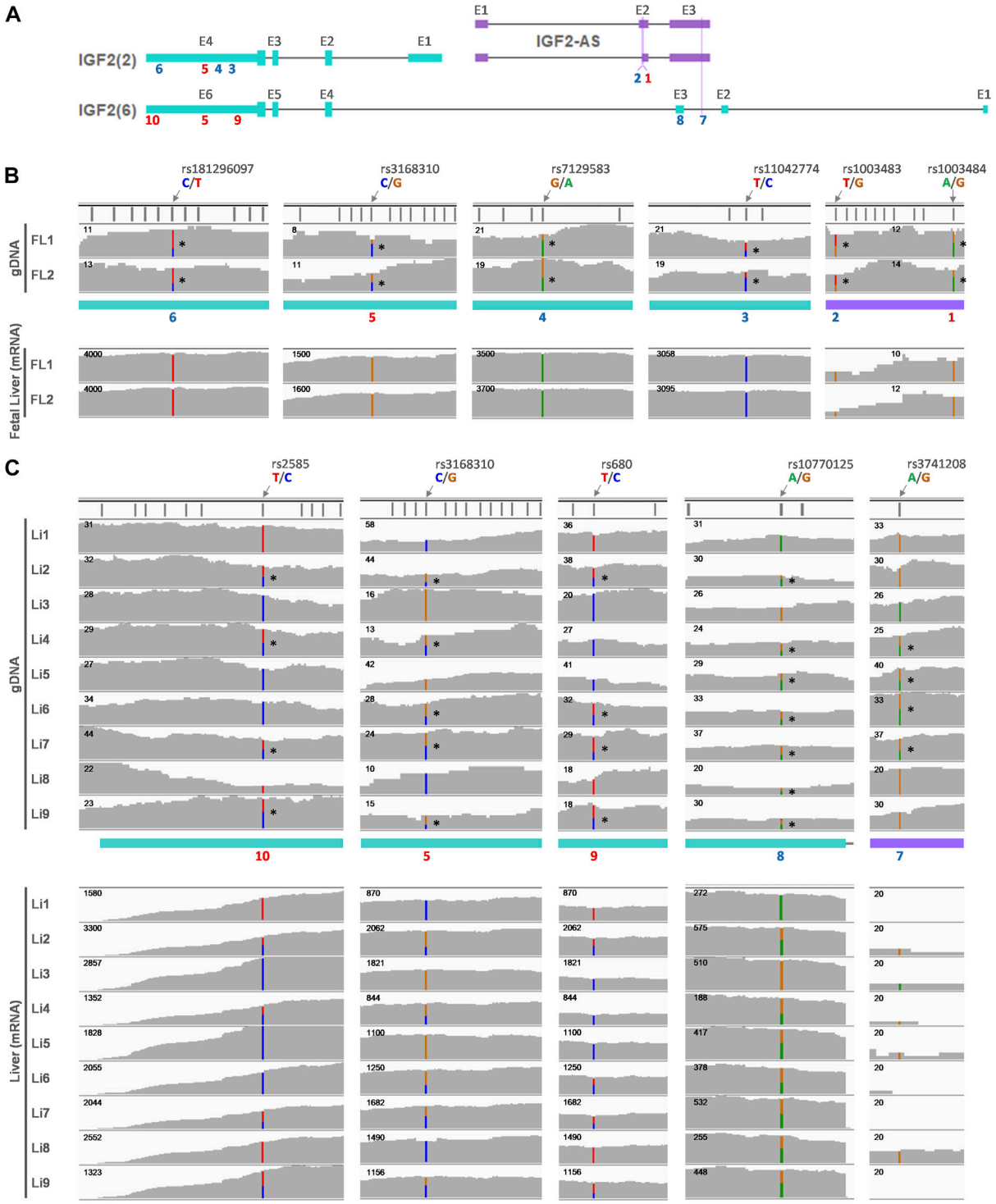


FIGURE 7 | Allelic expression of *IGF2* in fetal and adult human liver. **(A)** Predominant *IGF2* transcripts in either the fetal liver (FL) (*IGF2(2)*) or adult liver (Li) (*IGF2(6)*) are displayed along with antisense transcripts (*IGF2-AS*). SNP locations are marked with numbers on exons (E) and numbers in red (5, 9, and 10) indicate the same SNPs as in **Figure 7**. **(B)** Heterozygous alleles in the gDNA were identified using H3K4me1 ChIP-seq data retried from GSE63634 for fetal liver at 12 weeks of gestation and are denoted with stars (*) on the right side of SNPs. Corresponding sites were displayed below for fetal liver mRNA expression in individual-matched RNA-seq. **(C)** WGS data obtained from hum0158.v2 for normal adult liver, and heterozygous alleles are marked with stars (*) on the right side of SNPs. Corresponding mRNA expression in the adult liver derived from RNA-seq from the same dataset is shown below.

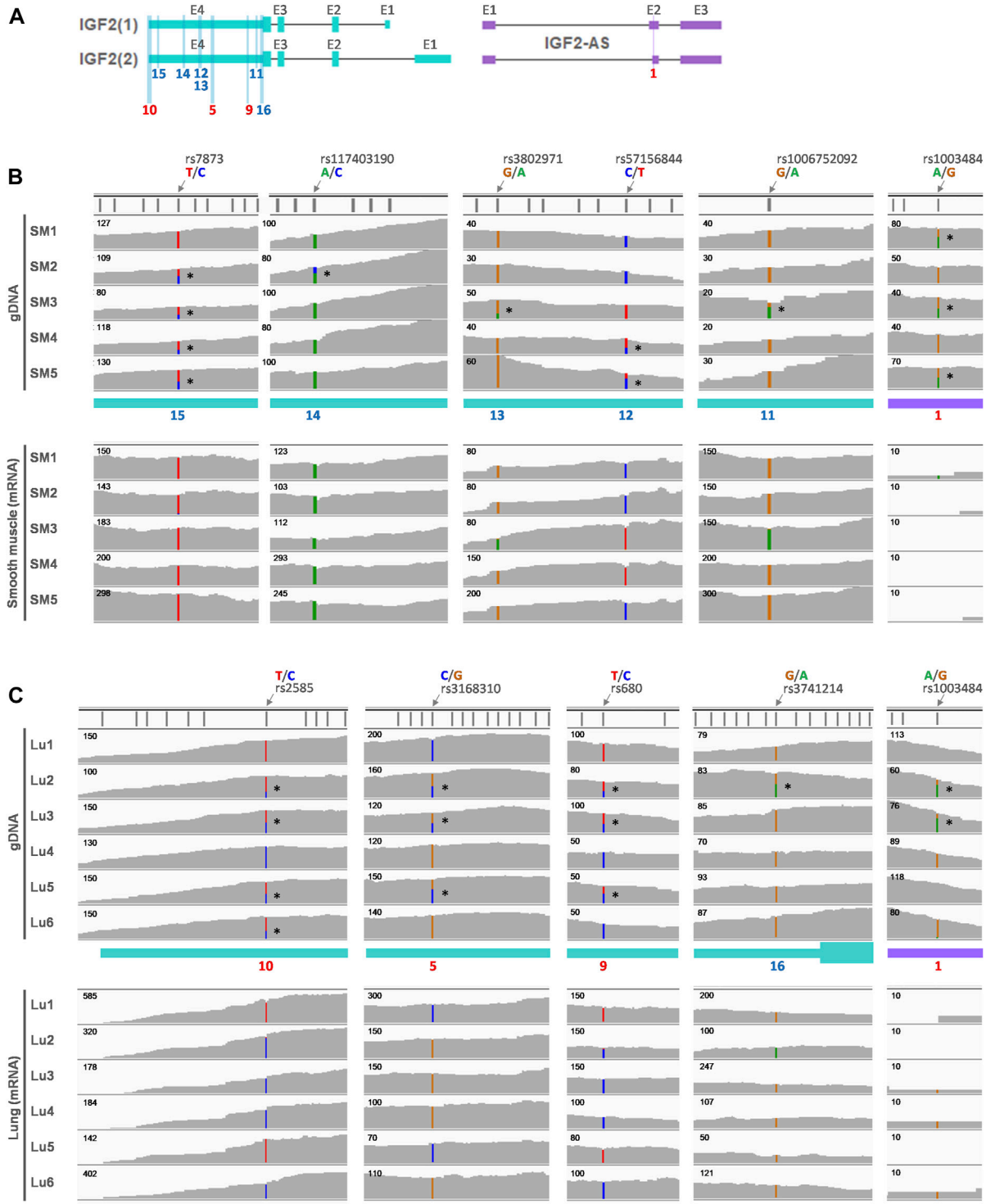


FIGURE 8 | Allelic expression of *IGF2* in adult human smooth muscle and lung. **(A)** Predominant *IGF2* transcripts (*IGF2(1)* and *IGF2(2)*) in both adult smooth muscle (SM) and adult lung (Lu) are displayed along with antisense transcripts (*IGF2-AS*). SNP locations are marked with numbers on exons (E) and numbers in red (5, 9, and 10) indicate the same SNPs as in **Figure 7**. **(B)** Whole-exome sequencing (WES) data (SRP163897) generated using adult smooth muscle were retrieved, and heterozygous alleles are marked on the gDNA. Corresponding matched RNA-seq (SRP166862) were processed and allelic expression is displayed below. **(C)** Heterozygous alleles were identified using WES data (PRJNA395106) of the human adult lung and marked with *. Corresponding matched RNA-seq from the same dataset (PRJNA395106) were used to show allelic expression.

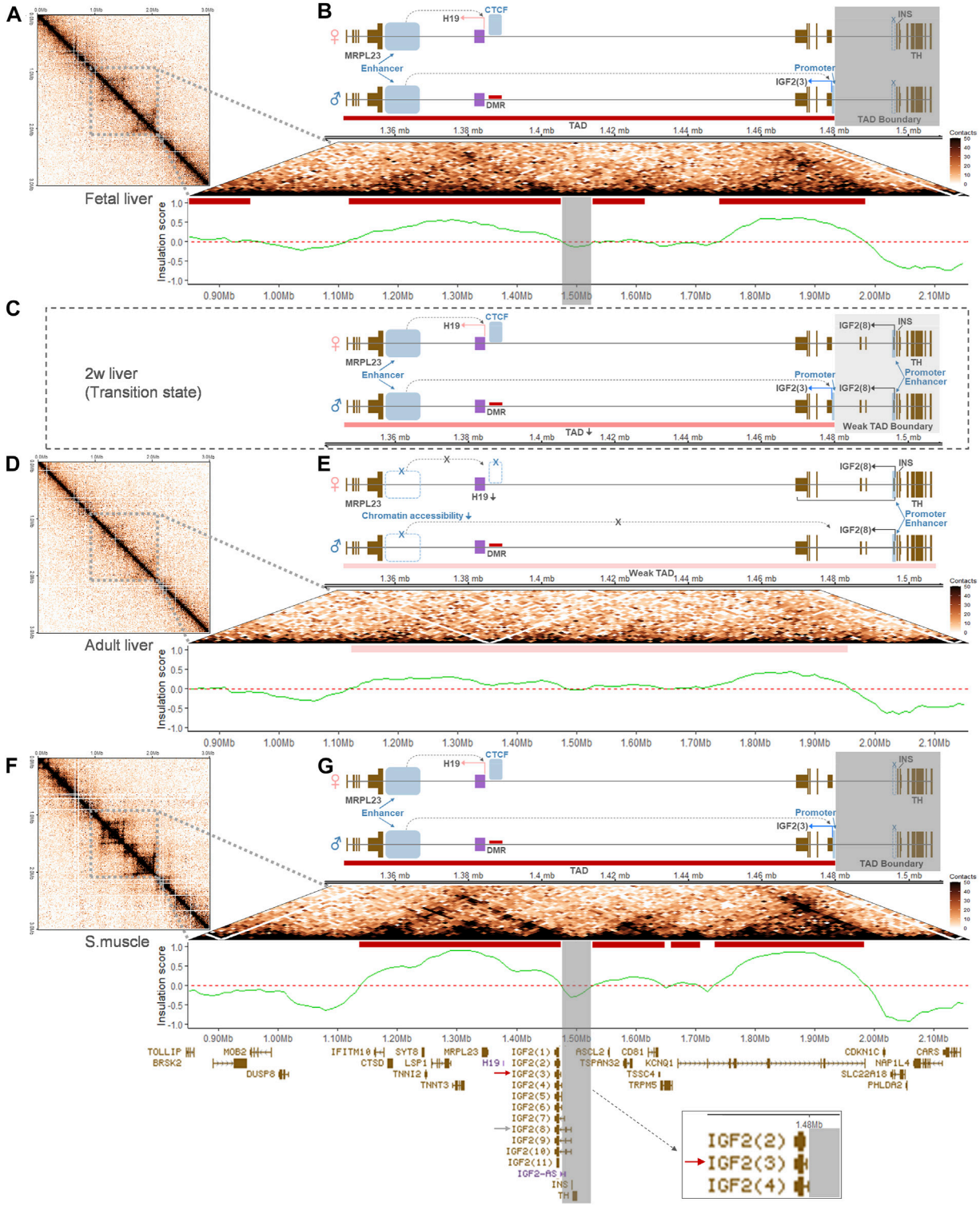


FIGURE 9 | Chromatin interactions and schematic models of gene regulation at the porcine *H19/IGF2* locus. For the chromosomal region (chr2:0-3000000) containing the *H19/IGF2* locus, two-dimensional heatmaps of Hi-C contact matrices were generated at a 10-kb resolution using the GENOVA R-package. Corresponding truncated pyramid plots zoomed in the locus. TADs were identified based on insulation scores and are indicated with red bars. **(A,B)** In the fetal liver, a TAD boundary is indicated with the grey perpendicular shades and a schematic diagram displays the long-range enhancer-promoter communication for paternal *IGF2(3)* expression within the TAD. Active regulatory elements are denoted with blue rectangles. In the maternal allele, CTCF binds to the hypomethylated region and expression of *H19* is promoted by an active enhancer. **(C)** A proposed schematic model of less chromatin interaction and week TAD boundary is shown for a transition

(*Continued*)

FIGURE 9 | state around 2 weeks (2w) for the porcine liver. Both paternal *IGF2(3)* expression and biallelic *IGF(8)* expression are indicated. **(D,E)** In the adult liver, a week TAD throughout the region is marked with pink bars. Also, low chromatin accessibility in the downstream of *H19* and altered gene regulation are represented in the diagram, along with biallelic *IGF(8)* expression indicated by black bent arrows on both alleles. **(F,G)** In skeletal muscle, the grey perpendicular shades denote a TAD boundary and paternal *IGF2(3)* expression is indicated, similarly to the ones in the fetal liver **(A,B)**. Gene transcripts expressed throughout the region are displayed in the bottom track. A red arrow marks the predominant *IGF2* isoform, *IGF2(3)*, in the fetal liver **(B)** and skeletal muscle **(G)**, and a grey arrow points at the long-form, *IGF2(8)*, which is expressed predominantly in the adult liver **(E)**. Both *IGF2(3)* and *IGF2(8)* are expressed in the proposed transition state of pigs **(C)**.

3.6 Topologically Associating Domains in the *H19/IGF2* Imprinted Cluster and Schematic Models Represent Insulation and Imprinting Boundaries

Hi-C datasets were analyzed to identify topologically associating domains (TADs) and TAD boundaries within and adjacent to the *H19/IGF2* locus. To determine whether Hi-C data from replicates of the liver and skeletal muscle can be merged, concordance of the data was estimated using GenomeDISCO (Ursu et al., 2018). For both fetal and adult livers, concordance scores for all pairwise comparisons of Hi-C matrices in multiple resolutions passed a threshold of 0.8 while also passing the recommended threshold of 0.8 at a 50-kb resolution (Ursu et al., 2018), although the scores tended to decrease in higher resolutions (**Supplementary Figure S9**). In addition, smoothed matrices produced from GenomeDISCO procedures displayed similarities within the tissue groups, but not between the groups (**Supplementary Figure S10–S15**). For skeletal muscle of 2-week-old pigs, concordance scores from pairwise comparison also passed the threshold and smoothed matrices were different from those of fetal and adult livers (**Supplementary Figures S16, S17**). Total number of trimmed paired-end reads of each group were similar: ~2.35 billion for fetal liver, ~2.90 billion for the liver, and ~2.24 billion for skeletal muscle (**Supplementary Table S7**). Based on high concordance and comparable amount of the data, matrices of three, three, and two replicates for each group (fetal liver, adult liver, and skeletal muscle), respectively, were merged to increase resolution. After merging, Hi-C matrices at a 10-kb resolution were visualized using two-dimensional heatmaps, and the matrices of the fetal liver and skeletal muscle exhibited stronger contact interactions between approximately 1.0 and 2.0 Mb of pig chromosome 2 than that of the adult liver (**Figure 9**). In the fetal liver and skeletal muscle, boundaries between TADs (i.e., TAD boundaries) encompassing the locus where the first exon of *IGF2(8)* starts, but not the first exon of *IGF2(3)*, was found (**Figures 9A,F**). In addition, muscle from fetal pigs contained a TAD boundary at the locus containing the first exon of *IGF2(8)* which was not overlapped with the first exon of *IGF2(3)* (**Supplementary Figure S18**). In contrast, in the adult liver, a week TAD throughout the region based on insulation scores was revealed, and TAD boundaries were absent at the *H19/IGF2* locus (**Figure 9D**). We expected lowered chromatin interaction in the liver at a transition state based on the enriched regulatory elements and gene expression patterns (**Figure 3**), but Hi-C data were absent (**Figure 9C**). In schematic diagrams, we propose chromatin reorganization in the liver that can weaken TAD and TAD boundaries and alter gene regulation, resulting in conversion of monoallelic expression of *IGF2(3)* in the fetal liver (**Figure 9B**) *via* both monoallelic expression of *IGF2(3)* and biallelic expression of *IGF2(8)* in the neonatal liver (**Figure 9C**) to biallelic expression of *IGF2(8)* in the adult liver (**Figure 9E**). In skeletal muscle, consistent *IGF2(3)* expression throughout the development and similar chromatin interaction frequencies between pre- and post-natal stages suggested robust imprinted gene regulation underlying monoallelic expression of the *IGF2* transcript (**Figure 9G**).

4 DISCUSSION

In this study, we present comprehensive imprinting status of the conserved paternally imprinted *H19/IGF2* cluster including developmentally regulated and tissue-specific allelic *IGF2* gene expression in the pig and human. By comparing methylome of parthenogenetic (diploid uni-maternal) embryos with bi-parental control embryos, while reducing genetic variability with triplicates of each sample, the porcine *H19* DMR was identified. Previously, the paternal methylation imprint on the *H19* germline DMR, which is fully methylated in sperm and unmethylated in oocytes, was reported in pigs in the form of a group of three DMRs (Park et al., 2009). On the paternal allele of the *H19* DMR, however, demethylation temporarily occurs and then it is remethylated by the morula stage (Park et al., 2009). In addition, differential expression of *IGF2* between androgenetic, parthenogenetic, and *in vitro* fertilized control embryos was previously observed from the blastocyst stage around day 10 (Park et al., 2011). Because, in this study, parthenogenetic and control embryos were recovered later at embryonic day 21at which the dynamic methylation changes were passed, the detected *H19* DMR could be consistent with the germline DMR between sperm and oocytes. The recovery day 21 was also before morphological degeneration of parthenogenetic embryos occurs at around day 30–35 (Bischoff et al., 2009; Hwang et al., 2020) so that we could prevent confounding effects other than genetic effects. On the other hand, putative *IGF2* DMRs, which were hypermethylated in sperm DNA of Swiss Landrace and Swiss Large White (Giannini and Braunschweig, 2009), were not found in the current study possibly due to breed-specific effects on DNA methylation as described previously (Hwang et al., 2020).

Integrative analyses of ATAC-seq, ChIP-seq, and RNA-seq datasets provide effective strategies to precisely and spatiotemporally elucidate epigenetic regulatory elements and their genetic variations that affect gene expression (Buenrostro et al., 2013; Lara-Astiaso et al., 2014; Floc'hlay et al., 2020). While variations on the regulatory DNA are often buffered and

compensated by other regulatory elements so that redundant regulatory signals might be present in the *H19/IGF2* locus (Floc'hlay et al., 2020), clear distinctions of regulatory layers between the liver and muscle tissues were identified (**Figure 2**). As the ATAC signals are substantially correlated with H3K27ac (Lara-Astiaso et al., 2014), co-occurrence of ATAC and H3K27ac downstream of *H19* suggested that active enhancers were established from a poised state concomitantly with formation of open chromatin sites upon developmental and signaling cues (Creyghton et al., 2010). Interestingly, this activation of enhancers occurred in skeletal muscle, but not in the liver, of 6-month-old pigs, which leads to recruitment of tissue-specific transcription factors and drives tissue-specific gene expression (Ong and Corces, 2011). In addition, in eukaryotes, H3K4me3 is associated with transcriptional activation on active promoters and typically restricted to narrow regions at the 5′ end of the gene body (Santos-Rosa et al., 2002; Schneider et al., 2004; Pena et al., 2006; Wysocka et al., 2006). H3K4me3 marks overlapping the CTCF signal immediately upstream of *H19* might represent the active *H19* promoter regulated by binding of the transcription factor CTCF at close range. Also, H3K4me3 marks overlapping the promoter regions of *IGF2* transcripts [i.e., *IGF2(3)* and *IGF2(8)*] might represent transcriptionally active *IGF2* promoters. At both proximal and distal regions of TSSs the H3K27ac signal can be found (Creyghton et al., 2010), and thus overlaps of H3K27ac with H3K4me3 at close proximity to the promoter regions of *IGF2* transcripts might represent both active enhancers and promoters. We found that these overlaps of H3K27ac and H3K4me3 were present near the first exons of both *IGF2(3)* and *IGF2(8)* transcripts in the liver of 2-week-old pigs, indicating distinct gene regulation during early post-natal liver development.

Although many imprinted genes have been studied in the fetal stage because of their relevance to fetal growth (Peters, 2014; Tian, 2014; Tucci et al., 2019), the current study revealed that the expressed transcript isoform of *IGF2* in muscle tissues might be stably maintained during development and its monoallelic expression was identified in post-natal stages suggesting its role in mature muscle. In the liver of fetal pigs, there was a lack of informative SNPs, but our parthenogenesis studies with whole embryos showed paternal monoallelic expression in the embryonic stage (**Figure 1B**). In the human, it has been reported that *IGF2* gene transcription is driven by multiple promoters in fetal and non-hepatic adult tissues (Holthuizen et al., 1993; Monk et al., 2006), but we showed that the major form in these tissues of humans was *IGF2(2)* which is orthologous to porcine *IGF2(3)*. On the other hand, the liver-specific promoter (P1) drives *IGF2* gene transcription in the adult liver (Holthuizen et al., 1993; Monk et al., 2006). The corresponding adult liver-specific transcript is not currently annotated in the NCBI Gene database (https://www.ncbi.nlm. nih.gov/gene/3481), but we revealed the expression of adult liver-specific *IGF2(6)* (**Supplementary Figure S7**) which was orthologous to the porcine long-form transcript [*IGF2(8)*]. The *IGF2* transcripts from human fetal tissues including the liver is paternally imprinted and monoallelic, but expression becomes biallelic in the adult liver (Kalscheuer et al., 1993). In addition to this biallelic conversion, we revealed that a relatively high *IGF2* expression occurred in the adult human liver compared to other analyzed tissues where monoallelic expression remains (**Figure 6** and **Supplementary Figure S6**). Our analyses using pigs support the biallelic conversion and alternative promoter usage that might occur gradually at post-natal ages while ages for initiation of the conversion might vary (**Figure 4**). These allelic expression patterns were verified based on individual-matched genomic DNA sequence data from WGS and mRNA sequence data from RNA-seq in both pigs and humans (**Figures 5**, **7**, and **8**) using informative SNPs found on genomic DNA that served as markers to confirm allelic imbalance of mRNA expression (Castel et al., 2015; Ahn et al., 2021b; a). We primarily examined SNPs in non-overlapping exons to identify and analyze allelic expression at the isoform level while there is a previous study relied on a marker in the last overlapping exon (Braunschweig et al., 2011). In contrast to the *INS-IGF2* read-through script whose expression is spatially regulated in pancreatic islets (Jian and Felsenfeld, 2021), the expression of *IGF2-AS* transcript has been shown to be developmentally regulated as its imprinted paternal expression is relatively high in fetal stages and decreased in adults in both pigs and humans (Okutsu et al., 2000; Braunschweig et al., 2004). Moreover, the notion that the expression of *IGF2-AS* in fetal stages interferes with overlapping *IGF2* (Braunschweig et al., 2004) might be supported by our findings: relatively high expression of *IGF2-AS* in the fetal liver coincided with negligible expression of the long-forms [porcine *IGF2(8)* and human *IGF2(6)*] and low expression of *IGF2-AS* in adults coincided with high expression of the long-forms. However, it is also expected that the antisense role of *IGF2-AS* might be limited in normal tissues due to its low expression compared with high expression of *IGF2,* although increased expression of *IGF2-AS* has been reported in Wilms tumors (Okutsu et al., 2000). Rather, the biallelic conversion in the liver might ensue changes in chromatin structure and regulatory elements as well as antisense expression as discussed below.

The differences in chromatin interaction between fetal and adult livers (and also between skeletal muscle and adult liver) suggested not only transcript conversion, but also chromatin remodeling might occur toward changes in gene regulatory elements and reduce long-range enhancer-promoter communication in the adult liver (**Figure 9** and **Supplementary Figure S18**). In particular, hemi-methylation at the *H19* DMR in the adult liver was reported indicating maintenance of the imprint (Braunschweig et al., 2011); however, compared to the fetal liver, chromatin interaction indicated by the self-interacting TADs became weaker in the adult liver of pigs. This lower interaction might be related to less activity of the distal enhancer for the long-range communication, and removal of TAD boundaries might lead to use of the proximal enhancer for the long *IGF2* transcript in the liver. Also, a linkage between less *H19* expression indicating the weak distal enhancer and expression of the biallelic *IGF2* transcript (Ohlsson et al., 1994) was consistently observed in the livers of both pigs and humans (**Supplementary Figures S2 and S6**). Additionally, in between the fetal and adult stages, there might be a transition state in the liver that is characterized by co-existence of TADs and a weak TAD boundary which is permissive to the proximal enhancer activity

(Gong et al., 2018). In contrast, in skeletal muscle, TADs and TAD boundaries in fetal stages appeared to maintain in post-natal ages. Whether these TADs and TAD boundaries are variable at the single cell level will need to be further investigated (Farabella and Marti-Renom, 2020; Luppino et al., 2020), but existence of TADs and TAD boundaries at the porcine *IGF2* locus was evident. Their significant changes and remodeling in the liver might contribute to liver-specific modifications of *IGF2* allelic expression patterns. Our presentation of the pig Hi-C fills the gap in mammalian genomics, but unfortunately, in the human, Hi-C data from liver (GSE58752) and muscle (GSE87112) tissues that we processed displayed a very low resolution for this relatively narrow range of the *H19/IGF2* locus. Based on our current study that advances our understanding on tissue-specific genomic imprinting in the *H19/IGF2* cluster, studies on other animal species using multi-omics data can further comparatively delineate the *H19/IGF2* locus. Also, gene annotations for porcine *H19* and *IGF2-AS* and human *IGF2(6)* need to be updated due to their lack in the NCBI Gene database (www.ncbi.nlm.nih.gov/gene). In this study, based on previous studies reported *H19* and *IGF2-AS* expression status in pigs (Li et al., 2008; Braunschweig et al., 2011) and *IGF2(6)* in humans (Holthuizen et al., 1993; Monk et al., 2006) as well as our alignment and sequencing results, these genes were analyzed to present the complete landscape of genomic imprinting.

5 CONCLUSION

Our integrative omics analyses of genome, epigenome, and transcriptome revealed a comprehensive imprinting status at the *H19/IGF2* locus in pigs in comparison with humans. The porcine *H19/IGF2* imprinting cluster represented a long-term influence of genomic imprinting in muscle tissues but not in the liver which might be similar to that of the orthologous human gene cluster. To the best of our knowledge, this is the first study that describes relatedness between mono- to biallelic conversion of *IGF2* and alternative promoter usage in reorganized chromatin in the liver of adult pigs. The current approaches can be applied in cross-tissue and cross-species analyses to elucidate epigenetic mechanisms that underlie tissue growth and development.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI GEO repository under accession number GSE195528. The publicly available datasets analyzed for this study are summarized in Materials and Methods and listed in **Supplementary Table S8**.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration (RDA) of Korea (approval number NIAS 2015-670).

AUTHOR CONTRIBUTIONS

JA, SH, and KL conceived this project. JA contributed to methodologies of next-generation sequencing and Hi-C, JL and D-HK conducted polymorphism-based sequencing. I-SH, I-CC, and M-RP contributed to parthenogenesis. JA and KL contributed to the analysis of omics data. JA wrote the manuscript. JA, D-HK, and KL reviewed and edited the manuscript. SH and KL contributed to the funding acquisition. All authors approved the final manuscript.

FUNDING

This work was partially supported by the United States Department of Agriculture National Institute of Food and Agriculture Hatch Grant (Project No. OHO01304).

ACKNOWLEDGMENTS

The National Bioscience Database Center (NBDC, https://humandbs.biosciencedbc.jp/) allowed our access to WGS and RNA-seq data of normal human liver. High performance computing was conducted using supercomputers at the Ohio Supercomputer Center (www.osc.edu). We are thankful to Michelle Milligan for proofreading this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.920641/full#supplementary-material

REFERENCES

Ahn, J., Hwang, I. S., Park, M. R., Cho, I. C., Hwang, S., and Lee, K. (2020a). *The Landscape of Genomic Imprinting at the Porcine SGCE/PEG10 Locus from Methylome and Transcriptome of Parthenogenetic Embryos*, 10, 4037–4047. doi:10.1534/g3.120.40142*G3 Genes|Genomes|Genetics (Bethesda)*

Ahn, J., Hwang, I. S., Park, M. R., Hwang, S., and Lee, K. (2021a). *Genomic Imprinting at the Porcine DIRAS3 Locus*, 11, 1315. doi:10.3390/ani11051315*Anim. (Basel)*

Ahn, J., Hwang, I. S., Park, M. R., Hwang, S., and Lee, K. (2021b). Genomic imprinting at the porcine PLAGL1 locus and the orthologous locus in the human. *Genes (Basel)* 12, 541. doi:10.3390/genes12040541

Ahn, J., Wu, H., Lee, J., Hwang, I. S., Yu, D., Ahn, J. S., et al. (2020c). Identification of a novel imprinted transcript in the porcine GNAS complex locus using

methylome and transcriptome of parthenogenetic fetuses. *Genes (Basel)* 11, E96. doi:10.3390/genes11010096

Ahn, J., Wu, H., Lee, J., Hwang, I. S., Yu, D., Ahn, J. S., et al. (2020b). Identification of a novel imprinted transcript in the porcine GNAS complex locus using methylome and transcriptome of parthenogenetic fetuses.. *Genes (Basel)* 11, E96. doi:10.3390/genes11010096

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461. doi:10.1038/nature12787

Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a018382. doi:10.1101/cshperspect.a018382

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., and Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev.* 7, 1663–1673. doi:10.1101/gad.7.9.1663

Bartolomei, M. S., Zemel, S., and Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155. doi:10.1038/351153a0

Beacon, T. H., Delcuve, G. P., Lopez, C., Nardocci, G., Kovalchuk, I., Van Wijnen, A. J., et al. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clin. Epigenetics* 13, 138. doi:10.1186/s13148-021-01126-1

Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482–485. doi:10.1038/35013100

Bischoff, S. R., Tsai, S., Hardison, N., Motsinger-Reif, A. A., Freking, B. A., Nonneman, D., et al. (2009). Characterization of conserved and nonconserved imprinted genes in swine. *Biol. Reprod.* 81, 906–920. doi:10.1095/biolreprod.109.078139

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

Braunschweig, M. H., Owczarek-Lipska, M., and Stahlberger-Saitbekova, N. (2011). Relationship of porcine IGF2 imprinting status to DNA methylation at the H19 DMD and the IGF2 DMRs 1 and 2. *BMC Genet.* 12, 47. doi:10.1186/1471-2156-12-47

Braunschweig, M. H., Van Laere, A. S., Buys, N., Andersson, L., and Andersson, G. (2004). IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3. *Genomics* 84, 1021–1029. doi:10.1016/j.ygeno.2004.09.006

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218. doi:10.1038/nmeth.2688

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

Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., and Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. *Nat. Protoc.* 1, 2353–2364. doi:10.1038/nprot.2006.324

Coppola, C. J., Ramaker, C. R., and Mendenhall, E. M. (2016). Identification and function of enhancers in the human genome. *Hum. Mol. Genet.* 25, R190–R197. doi:10.1093/hmg/ddw216

Crane, E., Bian, Q., Mccord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., et al. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523, 240–244. doi:10.1038/nature14450

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936. doi:10.1073/pnas.1016071107

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castello, A., Benitez, R., Fernandez, A. I., et al. (2019). Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition. *PLoS One* 14, e0220708. doi:10.1371/journal.pone.0220708

Dechiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859. doi:10.1016/0092-8674(91)90513-x

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

Farabella, I., and Marti-Renom, M. A. (2020). TADs without borders. *Nat. Genet.* 52, 752–753. doi:10.1038/s41588-020-0670-x

Ferguson-Smith, A. C., Cattanach, B. M., Barton, S. C., Beechey, C. V., and Surani, M. A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature* 351, 667–670. doi:10.1038/351667a0

Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 12, 565–575. doi:10.1038/nrg3032

Floc'hlay, S., Wong, E., Zhao, B., Viales, R. R., Thomas-Chollier, M., Thieffry, D., et al. (2020). Cis-acting variation is common across regulatory layers but is often buffered during embryonic development. *Genome Res.* 31, 211–224. doi:10.1101/gr.266338.120

George, J. W., Fan, H., Johnson, B., Carpenter, T. J., Foy, K. K., Chatterjee, A., et al. (2019). Integrated epigenome, exome, and transcriptome analyses reveal molecular subtypes and homeotic transformation in uterine fibroids. *Cell Rep.* 29, 4069–4085. doi:10.1016/j.celrep.2019.11.077

Giannini, P., and Braunschweig, M. (2009). DNA methylation patterns at the IGF2-H19 locus in sperm of Swiss Landrace and Swiss Large White boars. *J. Anim. Breed. Genet.* 126, 475–479. doi:10.1111/j.1439-0388.2009.00802.x

Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., et al. (2018). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* 9, 542. doi:10.1038/s41467-018-03017-1

Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. *Methods Mol. Biol.* 1418, 335–351. doi:10.1007/978-1-4939-3578-9_16

Haig, D. (2004). Genomic imprinting and kinship: How good is the evidence? *Annu. Rev. Genet.* 38, 553–585. doi:10.1146/annurev.genet.37.110801.142741

Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365, 764–767. doi:10.1038/365764a0

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., Tilghman, S. M., et al. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489. doi:10.1038/35013106

Hayward, B. E., Kamiya, M., Strain, L., Moran, V., Campbell, R., Hayashizaki, Y., et al. (1998a). The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10038–10043. doi:10.1073/pnas.95.17.10038

Hayward, B. E., Moran, V., Strain, L., and Bonthron, D. T. (1998b). Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15475–15480. doi:10.1073/pnas.95.26.15475

Holthuizen, P., Van Dijk, M. A., Rodenburg, R. J., Koonen-Reemst, A. M., and Sussenbach, J. S. (1993). Transcriptional regulation of the major promoters of the human IGF-II gene. *Mol. Reprod. Dev.* 35, 391–393. doi:10.1002/mrd.1080350412

Howe, F. S., Fischl, H., Murray, S. C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? *Bioessays* 39, 1–12. doi:10.1002/bies.201600095

Hwang, I. S., Park, M. R., Lee, H. S., Kwak, T. U., Son, H. Y., Kang, J. K., et al. (2020). *Developmental and Degenerative Characterization of Porcine Parthenogenetic Fetuses during Early Pregnancy*, 10.*Anim. (Basel)*. doi:10.3390/ani10040622

Jacob, K. J., Robinson, W. P., and Lefebvre, L. (2013). Beckwith-wiedemann and silver-russell syndromes: Opposite developmental imbalances in imprinted regulators of placental function and embryonic growth. *Clin. Genet.* 84, 326–334. doi:10.1111/cge.12143

Jian, X., and Felsenfeld, G. (2021). Large parental differences in chromatin organization in pancreatic beta cell line explaining diabetes susceptibility effects. *Nat. Commun.* 12, 4338. doi:10.1038/s41467-021-24635-2

Juhling, F., Kretzmer, H., Bernhart, S. H., Otto, C., Stadler, P. F., Hoffmann, S., et al. (2016). metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* 26, 256–262. doi:10.1101/gr.196394.115

Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H., and Ropers, H. H. (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat. Genet.* 5, 74–78. doi:10.1038/ng0993-74

Kern, C., Wang, Y., Xu, X., Pan, Z., Halstead, M., Chanthavixay, G., et al. (2021). Functional annotations of three domestic animal genomes provide vital

resources for comparative and agricultural research. *Nat. Commun.* 12, 1821. doi:10.1038/s41467-021-22100-8

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, D. J., Kim, D. H., Hwang, I. S., Kim, D. E., Kim, H. J., Kim, J. S., et al. (2017). Generation of alpha-1, 3-galactosyltransferase knocked-out transgenic cloned pigs with knocked-in five human genes. *Transgenic Res.* 26, 153–163. doi:10.1007/s11248-016-9979-8

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D. A., David, E., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949. doi:10.1126/science.1256271

Latos, P. A., Pauler, F. M., Koerner, M. V., Senergin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. doi:10.1126/science.1228110

Li, C., Bin, Y., Curchoe, C., Yang, L., Feng, D., Jiang, Q., et al. (2008). Genetic imprinting of H19 and IGF2 in domestic pigs (*Sus scrofa*). *Anim. Biotechnol.* 19, 22–27. doi:10.1080/10495390701758563

Li, F., Wang, D., Song, R., Cao, C., Zhang, Z., Wang, Y., et al. (2020). The asynchronous establishment of chromatin 3D architecture between *in vitro* fertilized and uniparental preimplantation pig embryos. *Genome Biol.* 21, 203. doi:10.1186/s13059-020-02095-z

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 3, 13033997.

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

Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., et al. (2017). Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res.* 27, 865–874. doi:10.1101/gr.207456.116

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi:10.1126/science.1181369

Lleres, D., Moindrot, B., Pathak, R., Piras, V., Matelot, M., Pignard, B., et al. (2019). CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.* 20, 272. doi:10.1186/s13059-019-1896-8

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

Lunney, J. K., Van Goor, A., Walker, K. E., Hailstock, T., Franklin, J., Dai, C., et al. (2021). Importance of the pig as a human biomedical model. *Sci. Transl. Med.* 13, eabd5758. doi:10.1126/scitranslmed.abd5758

Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., Yunker, R., et al. (2020). Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.* 52, 840–848. doi:10.1038/s41588-020-0647-9

Monk, D., Sanches, R., Arnaud, P., Apostolidou, S., Hills, F. A., Abu-Amero, S., et al. (2006). Imprinting of IGF2 P0 transcript and novel alternatively spliced INS-IGF2 isoforms show differences between mouse and human. *Hum. Mol. Genet.* 15, 1259–1269. doi:10.1093/hmg/ddl041

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet.* 7, 45–49. doi:10.1016/0168-9525(91)90230-N

Morrison, J., Koeman, J. M., Johnson, B. K., Foy, K. K., Beddows, I., Zhou, W., et al. (2021). Evaluation of whole-genome DNA methylation sequencing library preparation protocols. *Epigenetics Chromatin* 14, 28. doi:10.1186/s13072-021-00401-y

Nordin, M., Bergman, D., Halje, M., Engstrom, W., and Ward, A. (2014). Epigenetic regulation of the Igf2/H19 gene cluster. *Cell Prolif.* 47, 189–199. doi:10.1111/cpr.12106

Ohlsson, R., Hedborg, F., Holmgren, L., Walsh, C., and Ekstrom, T. J. (1994). Overlapping patterns of IGF2 and H19 expression during human development: Biallelic IGF2 expression correlates with a lack of H19 expression. *Development* 120, 361–368. doi:10.1242/dev.120.2.361

Okutsu, T., Kuroiwa, Y., Kagitani, F., Kai, M., Aisaka, K., Tsutsumi, O., et al. (2000). Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. *J. Biochem.* 127, 475–483. doi:10.1093/oxfordjournals.jbchem.a022630

Ong, C. T., and Corces, V. G. (2011). Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293. doi:10.1038/nrg2957

Park, C. H., Kim, H. S., Lee, S. G., and Lee, C. K. (2009). Methylation status of differentially methylated regions at Igf2/H19 locus in porcine gametes and preimplantation embryos. *Genomics* 93, 179–186. doi:10.1016/j.ygeno.2008.10.002

Park, C. H., Uh, K. J., Mulligan, B. P., Jeung, E. B., Hyun, S. H., Shin, T., et al. (2011). Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos. *PLoS One* 6, e22216. doi:10.1371/journal.pone.0022216

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

Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., et al. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442, 100–103. doi:10.1038/nature04814

Peters, J. (2014). The role of genomic imprinting in biology and disease: An expanding view. *Nat. Rev. Genet.* 15, 517–530. doi:10.1038/nrg3766

Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., et al. (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3830–3835. doi:10.1073/pnas.96.7.3830

Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J., et al. (2004). The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. *Nat. Genet.* 36, 818–826. doi:10.1038/ng1397

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., Wysocka, J., et al. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283. doi:10.1038/nature09692

Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi:10.1093/nar/gku365

Reik, W., and Walter, J. (2001). Genomic imprinting: Parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32. doi:10.1038/35047554

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., et al. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411. doi:10.1038/nature01080

Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., Kouzarides, T., et al. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* 6, 73–77. doi:10.1038/ncb1076

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., et al. (2015). HiC-pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x

Shemer, R., Birger, Y., Dean, W. L., Reik, W., Riggs, A. D., and Razin, A. (1996). Dynamic methylation adjustment and counting as part of imprinting mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6371–6376. doi:10.1073/pnas.93.13.6371

Shen, Y., Yue, F., Mccleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120. doi:10.1038/nature11243

Sun, F. L., Dean, W. L., Kelsey, G., Allen, N. D., and Reik, W. (1997). Transactivation of Igf2 in a mouse model of Beckwith-Wiedemann syndrome. *Nature* 389, 809–815. doi:10.1038/39797

Suntsova, M., Gaifullin, N., Allina, D., Reshetun, A., Li, X., Mendeleeva, L., et al. (2019). Atlas of RNA sequencing profiles for normal human tissues. *Sci. Data* 6, 36. doi:10.1038/s41597-019-0043-4

Tang, Q., Gu, Y., Zhou, X., Jin, L., Guan, J., Liu, R., et al. (2017). Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. *Gigascience* 6, 1–9. doi:10.1093/gigascience/gix105

Thorvaldsen, J. L., Duran, K. L., and Bartolomei, M. S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* 12, 3693–3702. doi:10.1101/gad.12.23.3693

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82. doi:10.1038/nature11232

Tian, X. C. (2014). Genomic imprinting in farm animals. *Annu. Rev. Anim. Biosci.* 2, 23–40. doi:10.1146/annurev-animal-022513-114144

Tian, X., Li, R., Fu, W., Li, Y., Wang, X., Li, M., et al. (2020). Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data. *Sci. China. Life Sci.* 63, 750–763. doi:10.1007/s11427-019-9551-7

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., and Erice Imprinting, G. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965. doi:10.1016/j.cell.2019.01.043

Ursu, O., Boley, N., Taranova, M., Wang, Y. X. R., Yardimci, G. G., Stafford Noble, W., et al. (2018). GenomeDISCO: A concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics* 34, 2701–2707. doi:10.1093/bioinformatics/bty164

Ushiki, A., Zhang, Y., Xiong, C., Zhao, J., Georgakopoulos-Soares, I., Kane, L., et al. (2021). Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. *Nat. Commun.* 12, 2282. doi:10.1038/s41467-021-22470-z

Van Der Weide, R. H., Van Den Brand, T., Haarhuis, J. H. I., Teunissen, H., Rowland, B. D., De Wit, E., et al. (2021). Hi-C analyses with GENOVA: A case study with cohesin variants. *Nar. Genom. Bioinform.* 3, lqab040. doi:10.1093/nargab/lqab040

Wiench, M., John, S., Baek, S., Johnson, T. A., Sung, M. H., Escobar, T., et al. (2011). DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* 30, 3028–3039. doi:10.1038/emboj.2011.210

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86–90. doi:10.1038/nature04815

Xu, J., Pope, S. D., Jazirehi, A. R., Attema, J. L., Papathanasiou, P., Watts, J. A., et al. (2007). Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12377–12382. doi:10.1073/pnas.0704579104

Yan, F., Powell, D. R., Curtis, D. J., and Wong, N. C. (2020). From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* 21, 22. doi:10.1186/s13059-020-1929-3

Yan, L., Guo, H., Hu, B., Li, R., Yong, J., Zhao, Y., et al. (2016). Epigenomic landscape of human fetal brain, heart, and liver. *J. Biol. Chem.* 291, 4386–4398. doi:10.1074/jbc.M115.672931

Yang, Y., Fan, X., Yan, J., Chen, M., Zhu, M., Tang, Y., et al. (2021). A comprehensive epigenome atlas reveals DNA methylation regulating skeletal muscle development. *Nucleic Acids Res.* 49, 1313–1329. doi:10.1093/nar/gkaa1203

Yoshimizu, T., Miroglio, A., Ripoche, M. A., Gabory, A., Vernucci, M., Riccio, A., et al. (2008). The H19 locus acts *in vivo* as a tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12417–12422. doi:10.1073/pnas.0801540105

Yuan, R., Zhang, J., Wang, Y., Zhu, X., Hu, S., Zeng, J., et al. (2021). Reorganization of chromatin architecture during prenatal development of porcine skeletal muscle. *DNA Res.* 28, dsab003. doi:10.1093/dnares/dsab003

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137

Zhang, Y., Zhang, L., Yue, J., Wei, X., Wang, L., Liu, X., et al. (2019). Genome-wide identification of RNA editing in seven porcine tissues by matched DNA and RNA high-throughput sequencing. *J. Anim. Sci. Biotechnol.* 10, 24. doi:10.1186/s40104-019-0326-9

Zhao, Y., Hou, Y., Xu, Y., Luan, Y., Zhou, H., Qi, X., et al. (2021). A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat. Commun.* 12, 2217. doi:10.1038/s41467-021-22448-x

Zhou, J., So, K. K., Li, Y., Li, Y., Yuan, J., Ding, Y., et al. (2019). Elevated H3K27ac in aged skeletal muscle leads to increase in extracellular matrix and fibrogenic conversion of muscle satellite cells. *Aging Cell* 18, e12996. doi:10.1111/acel.12996

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 26 July 2022
DOI 10.3389/fvets.2022.934728

OPEN ACCESS

EDITED BY
Lucas Lima Verardo,
Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

REVIEWED BY
Danielly Beraldo dos Santos Silva,
University of José do Rosário Vellano, Brazil
Sayed Haidar Abbas Raza,
Northwest A&F University, China

*CORRESPONDENCE
Hailong Zhou
zhouhl@hainanu.edu.cn
Tieshan Xu
xutieshan760412@163.com

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

RECEIVED 03 May 2022
ACCEPTED 27 June 2022
PUBLISHED 26 July 2022

CITATION
Gu L, Jiang Q, Chen Y, Zheng X, Zhou H and Xu T (2022) Transcriptome-wide study revealed m6A and miRNA regulation of embryonic breast muscle development in Wenchang chickens. *Front. Vet. Sci.* 9:934728. doi: 10.3389/fvets.2022.934728

Transcriptome-wide study revealed m6A and miRNA regulation of embryonic breast muscle development in Wenchang chickens

Lihong Gu[1], Qicheng Jiang[2], Youyi Chen[3], Xinli Zheng[1], Hailong Zhou[2]* and Tieshan Xu[4]*

[1]Institute of Animal Science and Veterinary Medicine, Hainan Academy of Agricultural Sciences, Haikou, China, [2]School of Life Science, Hainan University, Haikou, China, [3]Wuzhishan Animal Science and Veterinary Medicine and Fishery Service Center, Wuzhishan Agricultural and Rural Bureau, Wuzhishan, China, [4]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

N^6-Methyladenosine (m6A) modification has been shown to play important role in skeletal muscle development. Wenchang chickens are commonly used as a high-quality animal model in researching meat quality. However, there have been no previous reports regarding the profile of m6A and its function in the embryonic breast muscle development of Wenchang chickens. In this paper, we identified different developmental stages of breast muscle in Wenchang chickens and performed m6A sequencing and miRNA sequencing in the breast muscle of embryos. Embryo breast muscles were weighed and stained with hematoxylin–eosin after hatching. We found that myofibers grew fast on the 10^{th} day after hatching (E10) and seldom proliferated beyond the 19^{th} day after hatching (E19). A total of 6,774 differentially methylated genes (DMGs) were identified between E10 and E19. For RNA-seq data, we found 5,586 differentially expressed genes (DEGs). After overlapping DEGs and DMGs, we recorded 651 shared genes (DEMGs). Subsequently, we performed miRNA-seq analysis and obtained 495 differentially expressed miRNAs (DEMs). Then, we overlapped DEMGs and the target genes of DEMs and obtained 72 overlapped genes (called miRNA-m6A-genes in this study). GO and KEGG results showed DEMGs enriched in many muscle development-related pathways. Furthermore, we chose WNT7B, a key regulator of skeletal muscle development, to perform IGV visualization analysis and found that the m6A levels on the WNT7B gene between E10 and E19 were significantly different. In conclusion, we found that miRNAs, in conjunction with m6A modification, played a key role in the embryonic breast muscle development of Wenchang chickens. The results of this paper offer a theoretical basis for the study of m6A function in muscle development and fat deposition of Wenchang chickens.

KEYWORDS

m6A, Wenchang chickens, miRNA, breast muscle, muscle development, fat deposition

Introduction

Unlike DNA modification, RNA modification is more complex and diverse. More than 150 types of RNA modifications are currently known, most of which occur after transcription (1). RNA modifications have an important impact on the diversity of RNA structure, type, and function. N^6-Methyladenosine (m6A) modification is the most common post-transcriptional RNA modification in eukaryotes and the most common internal modification in messenger RNA (mRNA) (2).

m6A modification is driven by the m6A modified enzyme system, which can be divided into three categories: writers, erasers, and readers. Among them, the writers, also called m6A methyltransferases, include METTL3/14, WTAP, and KIAA1429; their main function is to catalyze m6A modification of adenylate on mRNA. The erasers are also called m6A demethylases; their main function is to demethylate the modification of bases that have undergone m6A modification. Currently, the main m6A demethylases include two members, obesity-related protein (FTO) (3) and Alk B homologous protein 5 (ALKBH5) (4). These two demethylases lead to a reversible change in m6A modification. The readers, also called m6A binding proteins, are mainly used to identify m6A modified bases and activate downstream regulatory pathways, such as RNA degradation and miRNA processing. The m6A readers mostly include YTH family proteins, such as YTHDC1, YTHDF1, and YTHDF2. YTHDC1 interacts with the splicing factors SRSF3 and SRSF10 to affect transcript expression (5). YTHDF1 can be complexed with translation initiation action, thereby promoting the translation of m6A modified RNA (6). YTHDF2 can mediate the degradation of m6A modified RNA (7).

Increasing numbers of research studies, based on m6A modification, have found it to play important regulatory role in many physiological processes. Hao et al. showed that METTL3-mediated m6A modification plays an important role in biological processes such as mouse embryo development (8). Yang et al. found that RNA m6A modification is involved in the regulation of circadian rhythms in the chicken hypothalamus under both basal and chronic stress conditions (9). m6A modification also has many functions in embryonic stem cell differentiation, development of the nervous and hematopoietic systems, myogenesis, zygote formation, and embryonic development (10). The regulation of m6A methylation is integral to muscle generation during embryonic development. Meanwhile, Wang et al. (11) found that FTO is required for myogenic differentiation and suggested that the FTO-mediated mTOR-PGC-1α-mitochondrial axis plays crucial role in myogenic differentiation. Zhang et al. (12) detected the specific expression of METTL3 in the vascular system of zebrafish. It is speculated that m6A modification is closely related to blood development, and the key role of m6A methylation modification in the development of hematopoietic stem cells has been revealed for the first time. m6A is also an important regulator of muscle development. Chen et. al. (13) explored the involvement of m6A mRNA modifications in mediating muscle regulation. Dang et al. (14) identified the key genes involved in cattle muscle growth and m6A modification development by bioinformatics analysis. The results showed that the differentially expressed genes modified by m6A are mainly involved in skeletal muscle contraction, steroid biosynthesis, redox process, the PPAR pathway, and fatty acid metabolism. Finally, Yang et al. (15) used m6A-seq to analyze bovine myoblasts and myotubes and found that m6A methylation was an abundant modification in mRNA. Furthermore, using experiments, they confirmed that four genes related to myogenesis exhibited differential changes in both m6A and mRNA levels during bovine myoblast differentiation, indicating that they can be potential candidate targets for m6A regulation of skeletal myogenesis.

MicroRNA (miRNA) is a type of noncoding RNA that exists widely in eukaryotes with a length of about 18–25 nt. miRNA plays an important role in the post-transcriptional regulation of target genes and is widely involved in various biological processes including growth and development, immunity, proliferation, and apoptosis. Shen et al. found that the expression of miR-152 can affect the quality of pork (16). Jebessa et. al. (17) investigated the underlying molecular mechanisms of skeletal muscle development based on differentially expressed genes and miRNAs. Meanwhile, Li et. al. (18) summarized miRNA related to muscle development, providing a better understanding of skeletal muscle development.

The rearing of Wenchang chickens is the most economically important livestock sector in Hainan province, with a 1.78 billion dollars output value in 2020. Wenchang chickens are famous for their rough feeding resistance and heat resistance and have a higher intramuscular fat and moderate subcutaneous fat content. Our research showed that the fastest stage of embryonic breast muscle development of Wenchang chickens was E10, and growth had slowed by E19. Given that m6A methylation modification plays an important role in skeletal muscle development, we speculated that modification of m6A methylation might be a crucial regulator of the growth rate of embryonic breast muscle. Herein, we performed m6A sequencing of breast muscle on embryos at E10 and E19. We also performed miRNA-seq to explore whether some genes were regulated by m6A and miRNA. We set out to determine the m6A profile and miRNA regulation in the embryonic breast muscle of Wenchang chickens; the results of this paper offer a basis for revealing the role of m6A modification in breast muscle development.

Abbreviations: DMGs, differentially methylated genes; DEGs, differentially expressed genes; DEMGs, shared genes between DEGs and DMGs; input, the data of RNA-seq; IP, the date of m6A-seq.

Methods

Ethics approval

This experiment was performed in accordance with animal welfare principles and was conducted under protocols approved by the Chinese Universities Union for the Protection of Animals.

All chickens were obtained from the Institute of Animal Science & Veterinary Medicine, Hainan Academy of Agricultural Sciences (IASVM-HAAS, Haikou, China). Ethics approval (reference number: IASVMHAAS-AE-202016) was conferred by the animal ethics committee of IASVM-HAAS, which is responsible for animal welfare. All experimental protocols were conducted in accordance with guidelines established by the Ministry of Science and Technology (Beijing, China).

Anatomy experiment

To identify different developmental stages of breast muscle in Wenchang chickens, we selected three eggs per day from the 8^{th} day after hatching (E8) and the 21^{st} day after hatching (E21). Embryo weight and breast muscle weight were recorded, and then, breast muscles from E8 to E21 were stripped from the bone and stained with HE staining by the same process as Gu et al. (19). Briefly, breast muscle samples of fourteen embryonic stages (E8–E21) were washed with running water and then dehydrated in a series of ethanol dilutions (75% for 4 h, 85% for 4 h, and 95% overnight) and then 100% ethanol for 2 h with two changes. Dehydrated tissues were treated with xylene three times and then embedded into paraffin blocks, trimmed, and cut to 4 μm. Paraffin ribbons were placed in a water bath at about 40 °C. Sections were mounted onto slides, air-dried for 30 min, and then dehydrated at 45 °C overnight. Sections were dewaxed with two changes of xylene for 10 min each and then hydrated with two changes of 100% ethanol for 3 min each, 95% and 80% ethanol for 1 min each, and, finally, rinsed in distilled water for 5 min. Slices were stained with hematoxylin and eosin (H&E). Sections from three samples of breast muscle were taken from each stage; five different fields were examined from each section, and pictures were taken under each field; cell counts were also performed for each field.

Sample collection

The Wenchang chicken embryos at E10 and E19 were purchased from the breeding farm of Hainan Chuanwei Wenchang Chickens Industry Co. Ltd. Three eggs were selected at E10 and E19; embryos were removed under aseptic conditions and stripped of their breast muscles; the left and right sides for each embryo were placed into different centrifuge tubes. The centrifuge tubes were placed immediately into liquid nitrogen and subsequently brought to the laboratory for storage at−80 °C for further use. The left and right breast muscles of chicken embryos were used for m6A-seq and RNA-seq analysis, respectively. In addition, another three embryonic breast muscle samples from E10 and E19 embryos were obtained and were used for miRNA-seq analysis.

RNA extraction and fragmentation

Total RNA was isolated and purified using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. We purified Poly (A) RNA from total RNA in three steps. The first step used NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA) to quantify the RNA and purity of each sample. The next step confirmed the RNA integrity, as assessed by Bioanalyzer 2100 (Agilent, CA, USA) with RIN number >7.0, using electrophoresis with denaturing agarose gel. The final step was purifying Poly (A) RNA from 50 μg total RNA using Dynabeads Oligo (dT) 25-61005 (Thermo Fisher, CA, USA) with two rounds of purification. The fragmentation buffer was added to the purified Poly (A) mRNA for fragmentation.

M6A immunoprecipitation, library construction, and sequencing

The fragmented RNA was divided into two parts. The first part was incubated for 2 h at 4 °C with m6A-specific antibody (No. 202003, Synaptic Systems, Germany) in IP buffer (50 mM Tris-HCl, 750 mM NaCl, and 0.5% IGEPAL CA-630). To create the cDNA, we made IP RNA reverse-transcribed with SuperScript™ II Reverse Transcriptase (Invitrogen, cat. 1896649, USA) according to the technical manual. Then, the products were used to synthesize U-labeled second-stranded DNAs with *E. coli* DNA polymerase I (NEB, cat.m0209, USA), RNase H (NEB, cat.m0297, USA), and dUTP Solution (Thermo Fisher, cat.R0133, USA). Because each adapter contains a T-base overhang, we added an A-base to the blunt ends of each strand for ligating the adapter. Single- or dual-index adapters were ligated to the fragments, and size selection was performed with AMPure XP beads. Subsequently, the heat-labile UDG enzyme (NEB, cat.m0280, USA) treatment of the U-labeled second-stranded DNAs was performed. The products were amplified with PCR under the following conditions: initial denaturation at 95 °C for 3 min; 8 cycles of denaturation at 98 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72 °C for 30 s; and then final extension at 72 °C for 5 min. The average insert size for the final cDNA library was 300 ± 50 bp.

The second part served as a control to construct a conventional transcriptome sequencing library directly. The two constructed sequencing libraries, m6A-seq Library (IP) and

RNA-seq Library (input), were separately subjected to high-throughput sequencing using Illumina NovaSeq™ 6000, and the sequencing mode was 150 PE.

M6A-seq and RNA-seq data filtering

M6A-seq technology employs a co-immunoprecipitation approach in which m6A-specific antibodies are incubated with RNA fragments that are randomly interrupted (IP) and fragments that grasp the m6A methylation modification are sequenced. One control group (input) was simultaneously sequenced in parallel and used to eliminate background during grasping with methylated fragments (20). So we performed m6A-seq in concert with RNA-seq. First, the reads containing adapter contamination, low-quality bases, and undetermined bases were removed by fastp (https://github.com/OpenGene/fastp) with default parameters, and valid data were then obtained. Sequence quality of IP and input samples were also verified using fastp. The R package exomePeak 1.8 (21) provided mapped reads for IP reads and input libraries, which identifies m6A peaks with bed or bigwig format. Furthermore, we use motif analysis software MEME 1.0 to identify motifs with high reliability in the peak area and record the width, E-value, PFM, PSSM of each motif, and its total position information in each peak sequence, then performed motif prediction for each group of samples, and analyzed the differential expression results.

M6A methylation peak screening and differential m6A analysis

The peak calling software and the R package exomePeak 1.8 were used to perform peak scanning on m6A samples and transcriptome samples to obtain the location of the peak on the genome, peak length, and the difference calculations between groups. These peaks were finally annotated using ChIPseeker 1.0.

Reference genome alignment

Valid data were aligned onto the reference genome of chicken (*Gallus gallus*, Version: V96) using the software HISAT2 1.0. According to the genomic annotation file gtf, number and distribution of alignment to reference genome reads, reads number, and proportions in exon and intron were counted. Valid data that enabled alignment to the reference genome, in accordance with the regional information of the reference genome, could be defined as alignment to exon (exonic), intron (intronic), and intergenic (intergenic spacer region) data.

Differentially expressed genes

StringTie 1.0 (22) was used to detect the expression levels for all mRNAs from Input libraries by calculating FPKM [total exon fragments /mapped reads (millions)×exon length (kB)]. The differentially expressed mRNAs were selected according to the criteria of $|\log2foldchange| \geq 1$ and $p < 0.05$ by R package edgeR 4.1 (https://bioconductor.org/packages/edgeR) (23).

Conjoint analysis of DEGs and DMGs

We overlapped DEGs and DMGs to explore gene modification by m6A and differential genes that were potentially connected with breast muscles. If a DEG overlapped with one or more differentially methylated m6A peaks, it was called a differentially methylated gene (DEMG). We examined up- and down-overlapping methylated m6A sites with up- and downregulated genes; we then obtained the upregulated genes with upregulated methylated m6A sites (hyper-up), the downregulated genes with upregulated methylated m6A sites (hyper-down), the upregulated genes with downregulated methylated m6A sites (hypo-up), and the downregulated genes with hypo-methylated m6A sites (hypo-down).

MiRNA library construction, sequencing, and analysis

The experimental flow was performed following standard procedures provided by Illumina. Small RNA sequencing libraries were prepared using the TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, USA). After library preparation was completed, the constructed libraries were sequenced using Illumina hiseq^{2500} with a sequencing read length of single-end 1×50 bp.

ACGT101-miR 4.2 (LC Sciences, Houston, Texas, USA) was used to analyze the miRNA data. First, 3'-adaptors and junk sequences were removed to obtain sequences with lengths in the range of 18–26 nt. The remaining sequences were then aligned (miRNA excluded) to the mRNA, Rfam, and Repbase databases and filtered. Valid data were then aligned to pre-miRNA and the genome for miRNA identification. The identified miRNAs were differentially analyzed, and the differential miRNAs were subjected to target gene prediction.

Integrated analysis of m6A-seq, MRNA-seq, and MiRNA-seq data

After differentially methylated m6A peaks and DEGs were obtained, we searched for DEGs to examine whether a

DEG overlapped with one or more differentially methylated m6A peaks. If such an overlap was demonstrated, it was termed a differentially methylated gene (DEMG). By the analysis of miRNAs, we obtained differential miRNAs (DEMs). The targets of DEMs were then predicted. Finally, we overlapped the targets of DEMs and DEMGs and obtained the shared genes.

Results

Determination of developmental regularity of Wenchang chicken breast muscle

We measured embryo weight and breast muscle weight and performed H&E staining for Wenchang chicken embryos at stages E8–E19. The weight of the breast muscle increased gradually from E8 to E17, (Figure 1A) then plateaued after E17, and decreased during the first few days of life, which may indicate that the breast muscle was initially in a proliferative differentiation phase; then during the first few days of life, the nutrition in the muscle was consumed, and the weight of the breast muscle decreased. By counting the results and section results in Figure 1B, it can be seen that the breast muscle cells were in the most proliferative period during E8–E12, and the myofibers gradually converged to form muscle bundles after E12; however, large numbers of free myocytes remained. By E17, the myofibers were progressively larger, free myocytes were progressively fewer, myocyte fusion events were gradually reduced, and the number of myofibers was concomitantly fixed. For E10–E12, the fastest growth took place at E10 and there was a gradual decline in number between E17 and E21 because of the progressive enlargement of myofibers with cessation of proliferation. Therefore, we selected days E10 and E19 to investigate the potential regulation of m6A modification in Wenchang chicken embryonic skeletal muscle using m6A-seq technology.

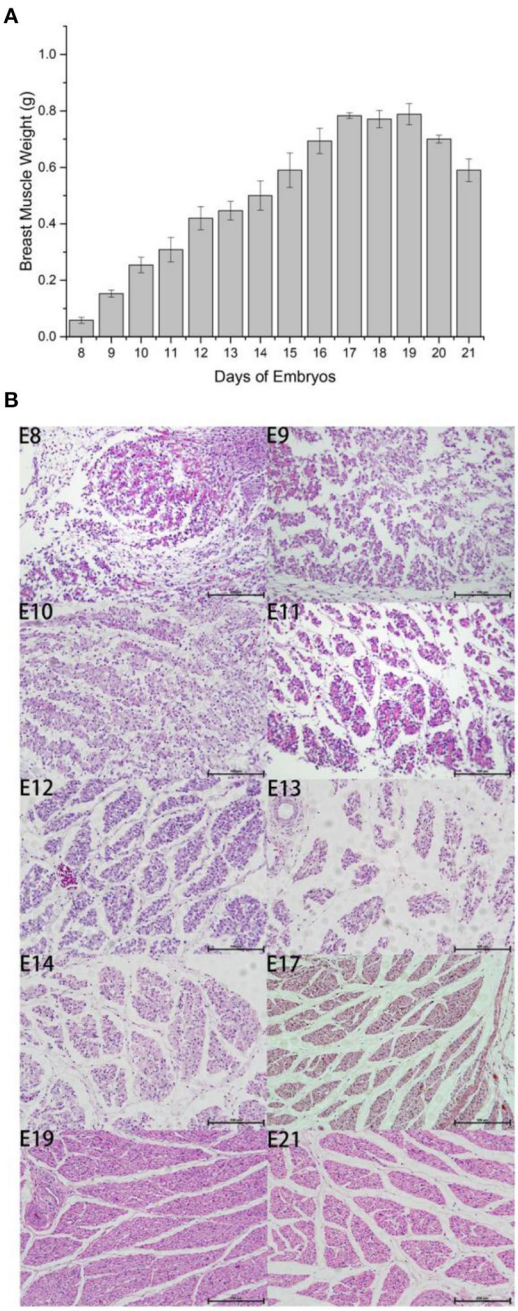


FIGURE 1
Outline of breast muscle development during the embryonic stages of Wenchang chickens. **(A)** Trend of breast muscle weights. **(B)** Embryonic breast muscle slices of Wenchang chickens. 8–14, 17, 19, and 21 represented to embryonic 8th day E8–E14, E17, E19, and E21, respectively.

Basic information of m6A-seq and RNA-seq

For RNA-seq (input), 60.53 G raw data were obtained from breast muscles at E10 and E19, with an average of 10.09 G per sample. The raw data were filtered by fastp, and a mean of 6.94 G was obtained per sample. For m6A-seq (IP), 64.07 G raw data were obtained at E10 and E19, with a mean of 10.67G per sample. After filtering, a total of 43.02 G of valid data were obtained, with a mean of 7.17 G per sample (Supplementary Table S1).

Alignment of valid data to the chicken reference genome

We then aligned the valid data to the chicken reference genome (Gallus gallus Version: V96) (Supplementary Table S2).

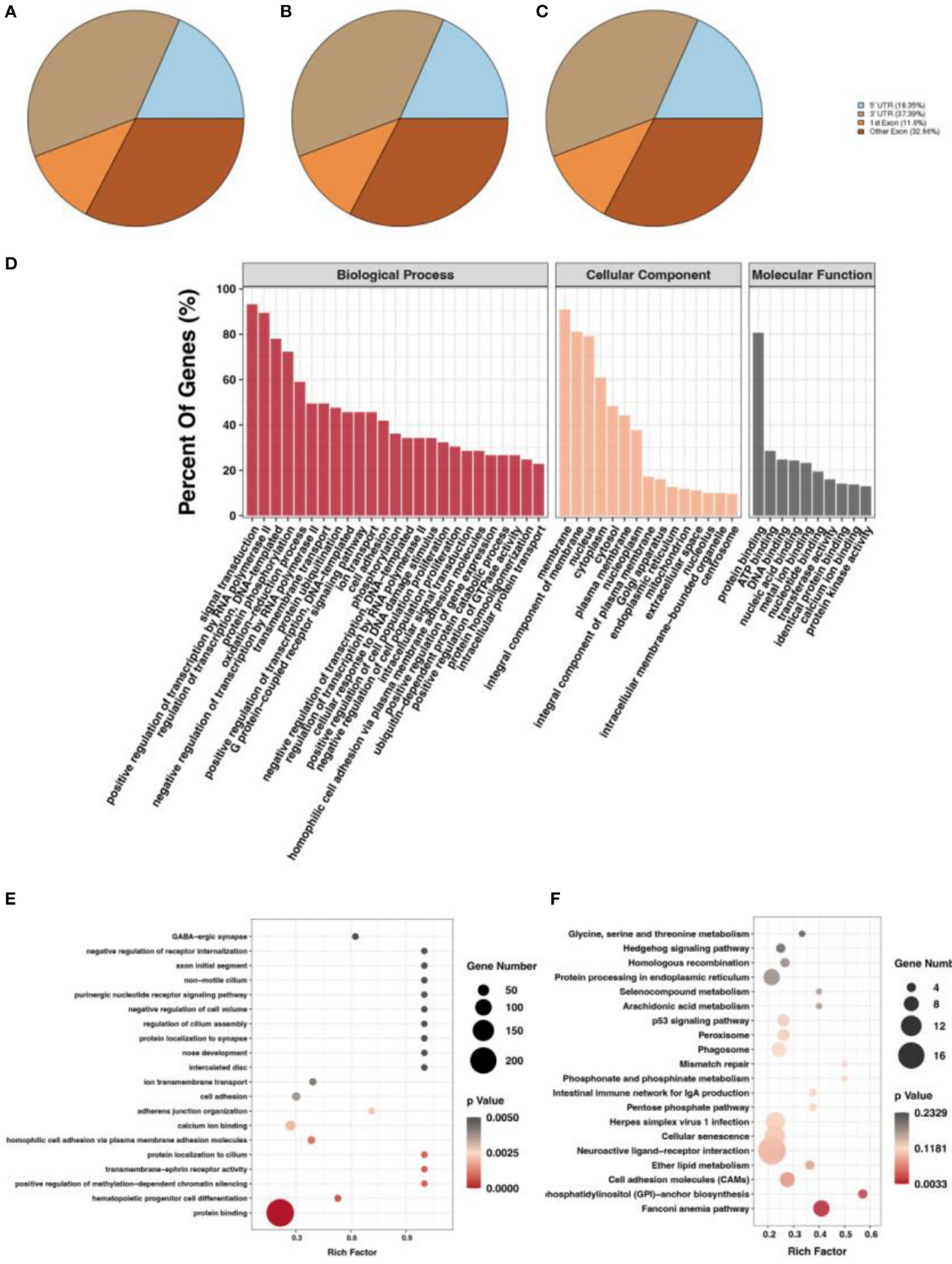


FIGURE 2
m6A methylation peak screening and differential m6A peak analysis. **(A)** m6A methylation peaks at E10. **(B)** m6A methylation peaks at E19. **(C)** Overlapped m6A methylation peaks between E10 and E19. **(D)** GO enrichment barplot of DMGs. **(E)** GO enrichment scatterplot of DMGs. **(F)** KEGG enrichment scatterplot of DMGs.

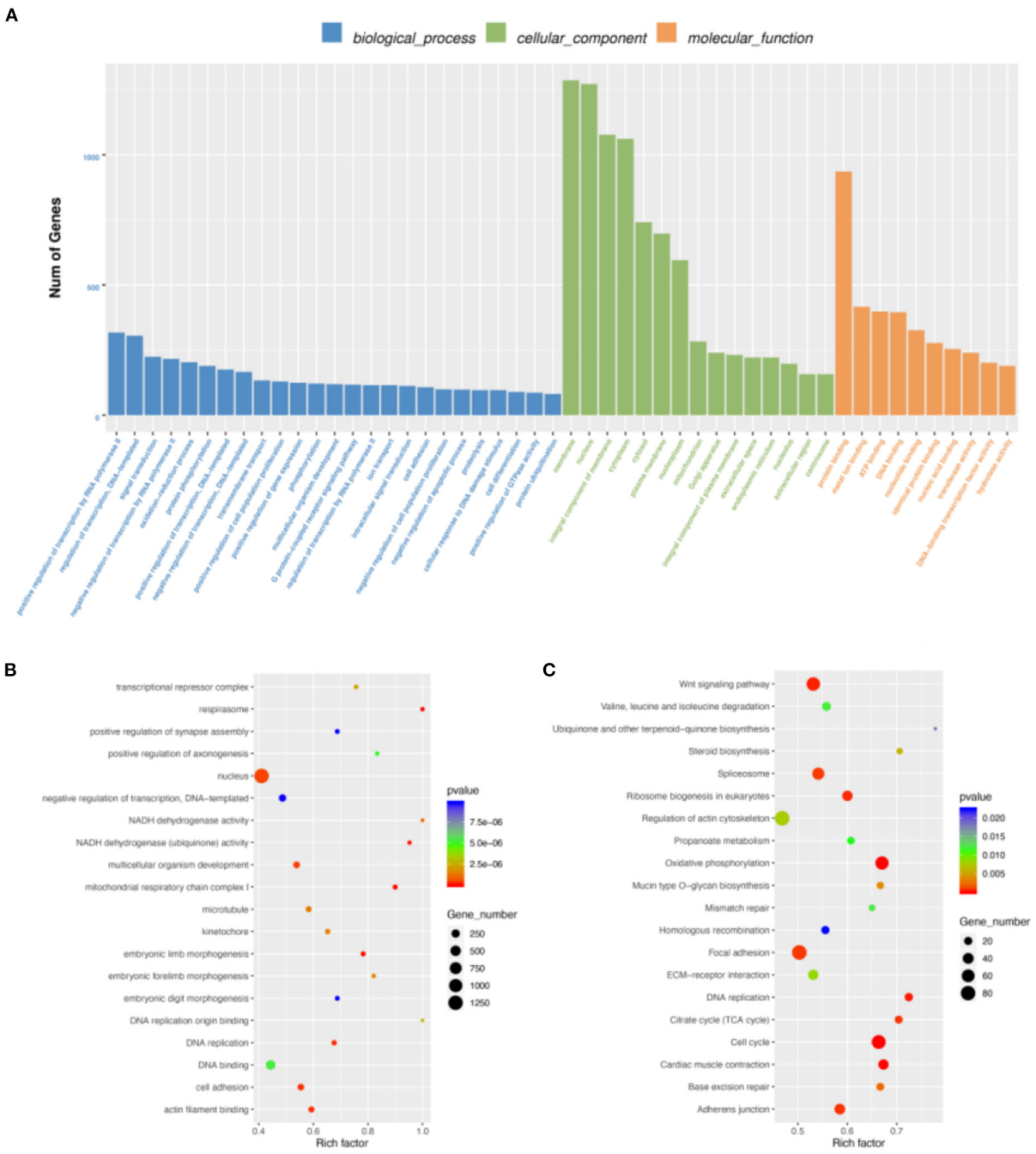


FIGURE 3
Differentially expressed gene (DEG) analysis. **(A)** GO enrichment barplot of DEGs. **(B)** GO enrichment scatterplot of DEGs. **(C)** KEGG enrichment scatterplot of DEGs.

Mean alignment rates were 90.83% for RNA-seq and 87.07% for m6A-seq. For m6A-seq, the majority of valid reads were aligned to exonic regions (exon) with 88.96, 87.7, and 87.67% for the three samples of E10 and 91.57, 92.96, and 92.49% for E19. Only a small proportion of the valid reads fell in intronic and intergenic regions. Similar to m6A-seq, most of the valid reads of RNA-seq aligned to exonic regions with 87.04, 87.81, and 89.07% for the three samples of E10 and 89.79, 90.51, and 89.84% for E19. In this study, most valid reads fell in exonic regions, which was consistent with the fact that the m6A-seq and RNA-seq libraries were constructed using Poly (A) RNA

10.3389/fvets.2022.934728

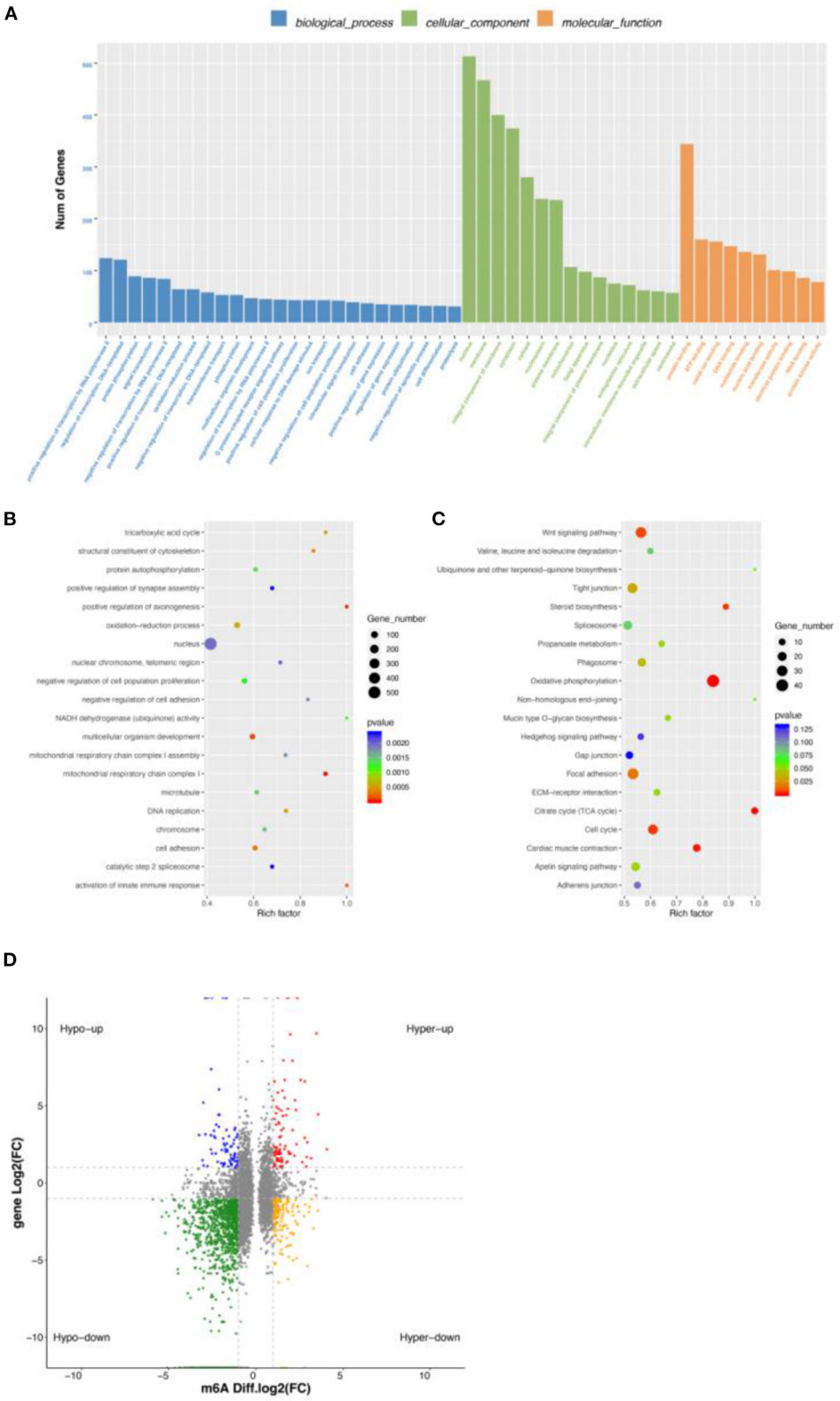


FIGURE 4
Conjoint analysis of DEGs and DMGs. **(A)** GO enrichment barplot of DEMGs. **(B)** GO enrichment scatterplot of DMGs. **(C)** KEGG enrichment scatterplot of DEMGs. **(D)** Regulatory relationship between m6A and differential genes.

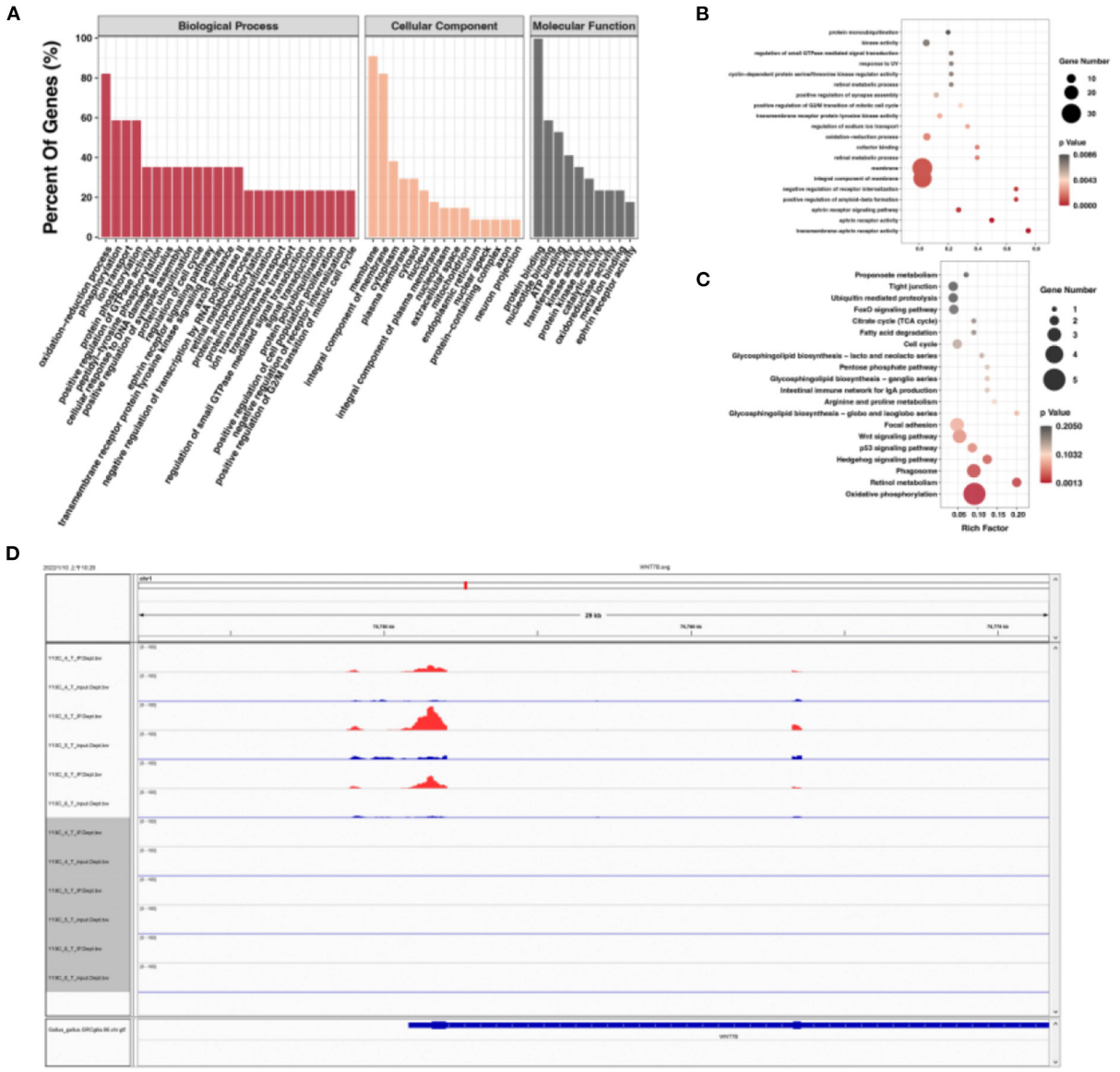


FIGURE 5
Integrated analysis of m6A-seq, mRNA-seq, and miRNA-seq data. **(A)** GO enrichment barplot of miRNA-m6A-genes. **(B)** GO enrichment scatterplot of miRNA-m6A-genes. **(C)** KEGG enrichment scatterplot of miRNA-m6A-genes. **(D)** The visualization of m6A abundances in *WNT7B* mRNA transcripts in E10 and E19.

(Supplementary Figure S1). The results above indicate that our sequencing results were accurate.

M6A methylation peak screening and differential m6A peak analysis

After screening for m6A methylation peaks, a total of 19,292 peaks were found at E10 (Supplementary Table S3), of which 8,538 were located in exons (44.26%), 7,213 in 3'UTR (37.39%), and only 3,541 in 5'UTR (18.35%; Figure 2A). In total, 11,294 genes were obtained with each of their transcripts overlapped with at least one m6A peak (called m6A genes in this study). For E19, a total of 16,843 peaks were found (Supplementary Table S4), of which 7,457 were located in exons (44.28%), 6,572 in 3'UTR (39.02%), and only 2,814 in 5'UTR (16.71%; Figure 2B). In total, 9,790 genes were obtained with each of their transcripts overlapped with at least one m6A peak. Differential analysis of the m6A methylated fragments from the two periods revealed that 6,774 differentially methylated peaks

existed, of which 1,514 were highly expressed at E10 and 5,260 were lowly expressed at E10; meanwhile, 2,517 differentially methylated peaks were located in exons (37.15%), 2,588 in the 3'UTR (38.20%), and 1,669 in the 5'UTR (24.64%) (Figure 2C). These m6A differentially methylated peaks overlapped with 5,565 genes (DMGs) (Supplementary Table S5).

To investigate the function of DMGs in breast muscle development, we performed GO and KEGG enrichment analyses. GO term displayed DMGs were most significantly enriched in protein binding in molecular function, signal transduction in biological process, and membrane in cellular component, as shown in Figure 2D. In addition, many GO terms related to muscle development, fat deposition, and m6A methylation were also significantly enriched, including branchiomeric skeletal muscle development, RNA N^6-methyladenosine methyltransferase complex, and fatty acid biosynthetic process (Figure 2E). KEGG results of DMGs are presented in Figure 2F. In total, we identified 123 pathways that were significantly enriched. Among them, the most significant pathways were neuroactive ligand–receptor interaction, cellular senescence, and cell adhesion molecules. Pathways related to muscle development and fat deposition were also significantly enriched, such as fatty acid elongation and phosphatidylinositol signaling system. We also identified the MYOG gene, a member of the MRFs (MyoD, MYOG, MyF5, and MRF4) family of myogenic regulators. The MYOG gene plays a key role in controlling the initiation of myoblast fusion, driving myoblast proliferation, and allowing the transition of mono-nucleated myoblasts into multinucleated myofibers (24).

Motif analysis

m6A methylation modification, as well as demethylation modification, begins with the binding of multiple binding proteins to a motif at which methylation sites occur. Motifs are essentially nucleic acid sequence patterns of biological significance. m6A methylation and demethylation enzymes can recognize the motifs and bind to them, thereby affecting gene expression and gene function. The identification of these motifs is important for mechanistic studies of gene expression regulation. To determine whether m6A peaks contain a motif of m6A by RRACH (i.e., R for purine, A for m6A, C for cytosine, and H for non-guanine bases), we performed motif predictions for E10 and E19 samples (Supplementary Figure S2) and found motifs in both E10 and E19.

Differentially expressed gene analysis

By DEG analysis of RNA-seq data, we obtained 5,586 DEGs (Supplementary Table S6). GO and KEGG analyses were then performed. Biological process of GO analysis showed that positive regulation of transcription by RNA polymerase II was the most enriched. Molecular function of GO analysis indicated protein binding, while membrane was the most enriched in cellular component. (Figure 3A). More important, we found many muscle-related GO terms and genes (Figure 3B), such as those in the MRF family (the MYF6 gene was more highly expressed at E19 than at E10 and the MYOG Gene was more highly expressed at E10 than at E19) and Pax family (Pax3 and Pax7 were both more highly expressed at E10 than E19). Pax3 and Pax7 in the Pax family have DNA activating transcription factor activity and bind RNA polymerase II specifically; they also play an important role in myofiber cell development (25). These results further suggest that some genes in the MRF family and Pax family may function during breast muscle development of Wenchang chickens at different embryonic ages. KEGG enrichment analysis of the DEGs showed that 160 pathways were significantly enriched including cell cycle, oxidative phosphorylation, cardiac muscle contraction, and DNA replication (Figure 3C). Among them, some muscle development-related pathways, such as the Wnt signaling pathway, were also obtained (26).

Conjoint analysis of DEGs and DMGs

To explore the relationship between m6A modification and gene expression, we carried out conjoint analysis of DEGs and DMGs. We overlapped DEGs and DMGs to produce 651 DEMGs in this study (Supplementary Table S7). GO functional analysis showed that most GO terms were enriched in cellular component (Figure 4A) and those DEMGs were significantly enriched in nucleus, positive regulation of transcription by RNA polymerase II, and protein binding (Figure 4B). KEGG results showed that 144 significantly enriched pathways were found, among which the most significant were Wnt signaling pathway, tight junction, and oxidative phosphorylation (Figure 4C). Interestingly, some muscle development-related pathways were also enriched, including the MAPK signaling pathway, an important regulatory pathway during myoblast differentiation (27). Activation of the MAPK signaling pathway can further increase skeletal myofiber cellular protein content, increase myofiber length and cross-sectional diameter, and make skeletal myofibers increase in mass when the number is unchanged (28).

We found a different correlation between methylated m6A level and gene expression abundance in E10 and E19 (Figure 4D). In 1,514 hyper-methylated m6A sites detected by m6A-seq, we found 314 genes with downregulated mRNA transcripts, that is, "hyper-down." There were 187 genes detected to have hyper-methylated m6A sites along with upregulated mRNA transcripts, that is, "hyper-up." In parallel to 5,260 hypo-methylated m6A sites, we found 214 genes with upregulated mRNA transcripts, that is "hypo-up." There were 1,279 genes examined to have hypo-methylated m6A sites along

with downregulated mRNA transcript, that is, "hypo-down." It is easy to see that hypo-down m6A level and gene expression in E10 compared with E19 were higher than in others, but in our previous research on ducks and geese, most genes were enriched in hypo-up and hyper-down indicating that m6A showed a negative correlation. This may be because those genes are not only regulated by m6A modification, but also by other regulatory factors such as miRNA.

MiRNA analysis

Some previous research demonstrated that m6A could influence miRNA production (29), so we performed miRNA-seq to explore the effect. In total, 150 significantly differentially expressed miRNAs (DEMs) were obtained by miRNA-seq analysis. Among them, 58 were upregulated and 92 were downregulated (Supplementary Figure S3). The target genes of DEMs were predicted, and 5,675 target genes were obtained. Interestingly, some m6A key enzyme genes were found to be the targets of DEMs. For example, METTL14 was found to be the target gene of gga-miR-6586-5p and gga-miR-132a-3p; and YTHDF1 was the target gene of gga-miR-6555-5p and gga-miR-1626-5p. This suggests that miRNA may be involved in regulating m6A modification.

Integrated analysis of m6A-seq, MRNA-seq, and MiRNA-seq data

Using overlapped DEMGs and the targets of DEMs, 72 miRNA-m6A-genes were obtained Supplementary Table S8. Then, GQ and KEGG functional analysis of miRNA-m6A-genes were performed. GO terms showed that they were significantly enriched in oxidation–reduction process, membrane, integral component of membrane, and cytoplasmic (Figure 5A). In addition, many GO terms were related to membrane, integral component of membrane, ephrin receptor activity, and others (Figure 5B). KEGG metabolic pathway enrichment analysis revealed that signal transduction, carbohydrate metabolism, lipid metabolism, glycan biosynthesis and metabolism, and amino acid metabolism were significantly enriched (Figure 5C). Among those miRNA-m6A-genes, we chose *WNT7B,* which is closely related with skeletal muscle development, to perform IGV visualization analysis and found significantly different m6A levels on the *WNT7B* gene between E10 and E19 (Figure 5D).

Discussion

There are many types of modifications on RNA, including m6A, N^1-methyladenosine modification (m1A), methylcytosine modification (m5C), and pseudouridine modification (Ψ). m6A was first discovered in the 1970s (30, 31); it is the most common internal modification in eukaryotic cell RNA and the best studied RNA modification. Studies have shown that m6A is common in mRNAs of different prokaryotes, eukaryotes, and viruses (32–34). m6A is the most common modification method in post-translational modification of eukaryotic mRNA, which accounts for more than 80% of RNA methylation (35). m6A is also a way for regulating mRNA metabolism and translation and plays important role in various physiological processes such as cell differentiation, development, and the stress response (2, 36).

m6A modification plays crucial role in many physiological processes such as myogenesis, embryonic development, and abnormal m6A modification can have huge effects on embryonic stem cell differentiation, zygote formation, and skeletal muscle development. Knockout of METTL3 in the embryonic stem cells before mouse embryo implantation will affect the stability of original pluripotency factor gene transcripts such as *Nanog* (homeobox protein NANOG), so that it cannot be downregulated, resulting in the obstruction of embryonic stem cell differentiation and the failure to fully terminate its naive state (37). Recently, another study has shown that the number of related gene transcripts cannot be changed during the development in oocytes of YTHDF2-messing mice; this phenomenon leads to infertility in this mouse (38). During embryonic development, muscle production is also inseparable from the regulation of m6A methylation. It has been shown that FTO as a demethylase of m6A can influence the differentiation of myoblasts. Deng et al. indicate that the FTO-mediated m6A modification in GADD45B mRNA drives skeletal muscle differentiation by activating the p38 MAPK pathway, which provides a molecular mechanism for the regulation of myogenesis via RNA methylation (39). Kudou et al. (40) found that METTL3 is involved in skeletal muscle differentiation of myogenic progenitor cells by mediating myogenic transcription factors such as MyoD, a key regulator of skeletal muscle differentiation. The results of our previous study illustrated the significance of m6A regulatory function in skeletal muscle development. Therefore, we selected breast muscles at E10 and E19 for preliminary experiments and found that m6A expression levels and methylation-modifying enzymes decreased with embryo age, indicating that m6A methylation modification played key role in the development of breast muscle in chicken embryos.

m6A methylation modification is a common phenomenon in post-transcriptional mRNA. Zhang et al. explored the expression profile of m6A in adult human tissues and detected a total of 101,340 methylation sites (41); the analysis of Dan et al. yielded 12,769 putative m6A sites within 6,990 coding gene transcripts and 250 noncoding ones (2). In this study, we found 18,068 peaks, which was consistent with the above studies. In addition, we found that the m6A methylation sites were mainly distributed in the 3'UTR regions, which are also the main binding sites of miRNAs. The m6A modification of the mRNA

was mostly enriched near the terminators and in the 3′-UTR (36); it showed the same mode of distribution as our results.

We obtained 5,565 DMG overlapped peaks between E10 and E19, indicating that m6A modification was prevalent in chicken gene modification. Using RNA-seq, we identified 5,586 DEGs between E10 and E19. We overlapped DMGs with DEGs and noted 651 DEMGs that were significantly affected by development stages and m6A modification. GO and KEGG results show that many genes were enriched in skeletal muscle development and fat deposition. Our results illustrated that m6A methylation modification had important effects on skeletal muscle development and fat deposition.

In addition, we found many muscle development-related genes and m6A modification-related genes, such as MYH gene, MYF5 gene, Pax gene, METTL14 gene, YTHDF1 gene, and MYOG gene among DEMGs; this indicated that both development stages and m6A modification might be regulators of skeletal muscle in Wenchang chickens. The roles of the myogenic regulatory factor (MRF) family (MyoD, myf5, myogenin, and MRF4) can cause the activation, proliferation, and differentiation of satellite cells (24). Studies have shown that Pax3 and Pax7 are upstream transcription factors of the MRF genes, which can cause expression of the MyoD gene in the embryonic period, thereby affecting the occurrence of muscle cells (25). Pax3 mutant mice will have missing limb muscles, but precursor cells and body muscle (including thoracic muscle) show normal differentiation (42); in murine knockouts of both Pax3 and Pax7, skeletal muscle development will end in the embryonic stage (43). After comparing gene expression levels in Wenchang chicken E10 and E19 breast muscle tissues, we found that the expression levels of the Pax7 gene, Pax3 gene, METTL14 gene, YTHDF1 gene, and MYOG gene were lower in E19 than in E10. However, the expression levels of MYF6 and MYH genes were higher in E19 than that in E10. Considering the fact that the embryonic breast muscle growth of Wenchang chickens is slower at E19 than at E10, Pax7, Pax3, METTL14, YTHDF1, and MYOG genes might be positive regulators, while MYF6 and MYH genes may be negative regulators.

miRNAs are important regulatory factors in skeletal muscle. To detect their effect on the development of breast muscle, we combined them with the results of m6A-seq data, to find those genes regulated by both m6A modification and miRNAs. We first carried out miRNA-seq analysis of embryonic breast muscles of Wenchang chickens at E10 and E19. We obtained 495 DEMs and 5,676 target genes of DEMs. Using overlapping the targets of DEMs with DEMGs, we obtained 72 miRNA-m6A-genes, which might be the regulators of embryonic breast muscle development of Wenchang chickens with their expression regulated by miRNAs and m6A modification. Among those miRNA-m6A-genes, we chose *WNT7B,* which is closely related with skeletal muscle development, to perform IGV visualization analysis and found significantly different m6A levels on the *WNT7B* gene between E10 and E19.

There may be another regulation mode here, that is, m6A modification regulates the expression of miRNA and then affects the expression of mRNA. The pri-miRNA is often modified by m6A methyltransferase and cleaved by methylated binding proteins to become pre-miRNA and eventually forms mature miRNA. Alarcon et al. found that METTL3 can methylate pri-miRNA, and the labeled pri-miRNA can be recognized and processed by DGCR8. By miRNA chip analysis, it was found that knockdown of METTL3 would reduce the binding between DGCR8 and pri-miRNA, reduce the expression of mature miRNA, and increase the content of unprocessed pri-miRNA. *In vitro* experiments, it has been confirmed that m6A could promote the processing of pri-miRNA. Finally, functional validation experiments revealed that METTL3 could promote miRNA maturation (44). Kyung-Won et al. uncovered the role of mRNA methylation on the abundance of AGO2 mRNA resulting in the repression of miRNA expression during the process of human aging (45). m6A modification can not only directly regulate gene expression, but also indirectly affect related regulatory factors to affect gene expression. We therefore suggest that m6A modification is important and extensive in organisms; however, the related regulatory mechanism remains to be studied.

Conclusion

In conclusion, we compared m6A modification profiles and gene expression levels at E10 and E19 of Wenchang chickens. We found that some genes shared by DMGs and DEMs in chicken embryonic breast muscle were enriched in skeletal muscle development-related pathways, indicating that m6A modification is one approach affecting skeletal muscle development and fat deposition. In addition, some target genes of DEMs overlapped with DEMGs, which indicated that miRNAs were potential regulators of m6A modification.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Sequence Read Archive (SRA); PRJNA836372 and PRJNA835987.

Ethics statement

The animal study was reviewed and approved by Institute of Animal Science & Veterinary Medicine, Hainan Academy of Agricultural Sciences.

Author contributions

LG, TX, and HZ conceived and designed the experiments. LG, YC, XZ, and TX performed the experiments. TX and QJ analyzed the data. TX, LG, and QJ wrote the paper. All authors read and approved the final manuscript.

Funding

This study was supported in part by the Key R&D projects in Hainan Province (SQ2020XDNY0161), the National Key R&D Program of China (Grant No. 2021YFD1300100), and the Project of Wenchang Chickens Research Institute (2,426), Basal Research Fund for Hainan Academy of Agricultural Sciences (JBKYYMF-2020-09).

Acknowledgments

We thank Dr. Yongzheng Huang for the data analysis and guidance. We thank Peng Yang for draft writing. We thank Shunshi Zhang for his cooperation in drawing. We thank Yifan Wen and Zijie Xu for sample collection.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.934728/full#supplementary-material

SUPPLEMENTARY TABLE S1
Basic information of sequenced data.

SUPPLEMENTARY TABLE S2
Mapping summary of chicken reference genome.

SUPPLEMENTARY FIGURE S1
Peak distribution of Wenchang chickens in E10 and E19.

SUPPLEMENTARY TABLE S3
m6A methylation peaks in E10.

SUPPLEMENTARY TABLE S4
m6A methylation peaks in E10.

SUPPLEMENTARY TABLE S5
Differentially methylated genes (DMGs) in E10 and E19.

SUPPLEMENTARY FIGURE S2
Motif sequence of m6A contained.

SUPPLEMENTARY TABLE S6
Differentially expressed genes (DEGs) in E10 and E19.

SUPPLEMENTARY TABLE S7
Shared genes by overlapped DEGs and DMGs, (DEMGs) in E10 and E19.

SUPPLEMENTARY FIGURE S3
Significantly differentially expressed miRNAs (DEMs) in E10 and E19.

SUPPLEMENTARY TABLE S8
miRNA-m6A-genes in E10 and E19.

References

1. Lui L, Lowe T. Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem.* (2013) 54:53–77. doi: 10.1042/bse0540053

2. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature.* (2012) 485:201–6. doi: 10.1038/nature11112

3. Zou S, Toh JD, Wong KH, Gao YG, Hong W, Woon EC. N^6-Methyladenosine: a conformational marker that regulates the substrate specificity of human demethylases FTO and ALKBH5. *Sci Rep.* (2016) 6:25677. doi: 10.1038/srep25677

4. Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell.* (2013) 49:18–29. doi: 10.1016/j.molcel.2012.10.015

5. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, et al. Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol Cell.* (2016) 61:507–19. doi: 10.1016/j.molcel.2016.01.012

6. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, et al. N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell.* (2015) 161:1388–99. doi: 10.1016/j.cell.2015.05.014

7. Chai RC, Chang YZ, Chang X, Pang B, An SY, Zhang KN, et al. YTHDF2 facilitates UBXN1 mRNA decay by recognizing METTL3-mediated m(6)A modification to activate NF-kappaB and promote the malignant progression of glioma. *J Hematol Oncol.* (2021) 14:109. doi: 10.1186/s13045-021-01124-z

8. Hao J, Xianfeng Y, Gao W, Wei J, Qi M, Han L, et al. The perturbed expression of m6A in parthenogenetic mouse embryos. *Genet Mol Biol.* (2019) 42:666–70. doi: 10.1590/1678-4685-gmb-2018-0212

9. Yang Y, Han W, Zhang A, Zhao M, Cong W, Jia Y, et al. Chronic corticosterone disrupts the circadian rhythm of CRH expression and m(6)A RNA methylation in the chicken hypothalamus. *J Anim Sci Biotechnol.* (2022) 13:29. doi: 10.1186/s40104-022-00677-4

10. Mendel M, Chen KM, Homolka D, Gos P, Pandey RR, McCarthy AA, et al. Methylation of structured RNA by the m(6)A writer METTL16 is essential for mouse embryonic development. *Mol Cell.* (2018) 71:986–1000 e1011. doi: 10.1016/j.molcel.2018.08.004

11. Wang X, Huang N, Yang M, Wei D, Tai H, Han X, et al. FTO is required for myogenesis by positively regulating mTOR-PGC-1alpha pathway-mediated mitochondria biogenesis. *Cell Death Dis.* (2017) 8:e.2702. doi: 10.1038/cddis.2017.122

12. Zhang C, Chen Y, Sun B, Wang L, Yang Y, Ma D, et al. m(6)A modulates haematopoietic stem and progenitor cell specification. *Nature.* (2017) 549:273–6. doi: 10.1038/nature23883

13. Chen W, Chen Y, Wu R, Guo G, Liu Y, Zeng B, et al. DHA alleviates diet-induced skeletal muscle fiber remodeling *via* FTO/m(6)A/DDIT4/PGC1alpha signaling. *BMC Biol.* (2022) 20:39. doi: 10.1186/s12915-022-01239-w

14. Dang Y, Dong Q, Wu B, Yang S, Sun J, Cui G, et al. Global landscape of m6A methylation of differently expressed genes in muscle tissue of liaoyu white cattle and simmental cattle. *Front Cell Dev Biol.* (2022) 10:840513. doi: 10.3389/fcell.2022.840513

15. Yang X, Wang J, Ma X, Du J, Mei C, Zan L. Transcriptome-wide N (6)-methyladenosine methylome profiling reveals m(6)A regulation of skeletal myoblast differentiation in cattle (Bos taurus). *Front Cell Dev Biol.* (2021) 9:785380. doi: 10.3389/fcell.2021.785380

16. Shen L, Gan M, Chen L, Zhao Y, Niu L, Tang G, et al. miR-152 targets pyruvate kinase to regulate the glycolytic activity of pig skeletal muscles and affects pork quality. *Meat Sci.* (2022) 185:108707. doi: 10.1016/j.meatsci.2021.108707

17. Jebessa E, Ouyang H, Abdalla BA Li Z, Abdullahi AY, Liu Q, et al. Characterization of miRNA and their target gene during chicken embryo skeletal muscle development. *Oncotarget.* (2018) 9:17309–24. doi: 10.18632/oncotarget.22457

18. Li XY, Fu LL, Cheng HJ, Zhao SH. Advances on microRNA in regulating mammalian skeletal muscle development. *Yi Chuan.* (2017) 39:1046–53. doi: 10.16288/j.yczz.17-112

19. Gu LH, Xu TS, Huang W, Xie M, Shi WB, Sun SD, et al. Developmental characteristics of pectoralis muscle in Pekin duck embryos. *Genet Mol Res.* (2013) 12:6733–42. doi: 10.4238/2013.December.13.6

20. Raza SHA, Khan R, Cheng G, Long F, Bing S, Easa AA, et al. RNA-Seq reveals the potential molecular mechanisms of bovine KLF6 gene in the regulation of adipogenesis. *Int J Biol Macromol.* (2022) 195:198–206. doi: 10.1016/j.ijbiomac.2021.11.202

21. Meng J, Lu Z, Liu H, Zhang L, Zhang S, Chen Y, et al. A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods.* (2014) 69:274–81. doi: 10.1016/j.ymeth.2014.06.008

22. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290–5. doi: 10.1038/nbt.3122

23. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

24. Wang Y, Schnegelsberg PN, Dausman J, Jaenisch R. Functional redundancy of the muscle-specific transcription factors Myf5 and myogenin. *Nature.* (1996) 379:823–5. doi: 10.1038/379823a0

25. Buckingham M, Relaix F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu Rev Cell Dev Biol.* (2007) 23:645–73. doi: 10.1146/annurev.cellbio.23.090506.123438

26. LeBoeuf B, Chen X, Garcia LR. WNT regulates programmed muscle remodeling through PLC-beta and calcineurin in caenorhabditis elegans males. *Development*. (2020) 147:dev181305. doi: 10.1242/dev.181305

27. Yin H, Zhao J, He H, Chen Y, Wang Y, Li D, et al. Gga-miR-3525 targets PDLIM3 through the MAPK signaling pathway to regulate the proliferation and differentiation of skeletal muscle satellite cells. *Int J Mol Sci.* (2020) 21:5573. doi: 10.3390/ijms21155573

28. Keren A, Tamir Y, Bengal E. The p38 MAPK signaling pathway: a major regulator of skeletal muscle development. *Mol Cell Endocrinol.* (2006) 252:224–30. doi: 10.1016/j.mce.2006.03.017

29. Diao LT, Xie SJ, Lei H, Qiu XS, Huang MC, Tao S, et al. METTL3 regulates skeletal muscle specific miRNAs at both transcriptional and post-transcriptional levels. *Biochem Biophys Res Commun.* (2021) 552:52–8. doi: 10.1016/j.bbrc.2021.03.035

30. Perry RP, Kelley DE, LaTorre J. Synthesis and turnover of nuclear and cytoplasmic polyadenylic acid in mouse L cells. *J Mol Biol.* (1974) 82:315–31. doi: 10.1016/0022-2836(74)90593-2

31. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from novikoff hepatoma cells. *Proc Natl Acad Sci U S A.* (1974) 71:3971–5. doi: 10.1073/pnas.71.10.3971

32. Beemon K, Keith J. Localization of N6-methyladenosine in the rous sarcoma virus genome. *J Mol Biol.* (1977) 113:165–79. doi: 10.1016/0022-2836(77)90047-X

33. Dubin DT, Taylor RH. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.* (1975) 2:1653–68. doi: 10.1093/nar/2.10.1653

34. Furuichi Y, Miura K, A. blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature.* (1975) 253:374–5. doi: 10.1038/253374a0

35. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA.* (1997) 3:1 233–47.

36. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell.* (2012) 149:1635–46. doi: 10.1016/j.cell.2012.05.003

37. Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, et al. m^6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science.* (2015) 347:1002–6. doi: 10.1126/science.1261417

38. Ivanova I, Much C, Di Giacomo M, Azzi C, Morgan M, Moreira PN, et al. The RNA m(6)A reader YTHDF2 is essential for the post-transcriptional regulation of the maternal transcriptome and oocyte competence. *Mol Cell.* (2017) 67:1059–67 e1054. doi: 10.1016/j.molcel.2017.08.003

39. Deng K, Fan Y, Liang Y, Cai Y, Zhang G, Deng M, et al. FTO-mediated demethylation of GADD45B promotes myogenesis through the activation of p38 MAPK pathway. *Mol Ther Nucleic Acids.* (2021) 26:34–48. doi: 10.1016/j.omtn.2021.06.013

40. Kudou K, Komatsu T, Nogami J, Maehara K, Harada A, Saeki H, et al. The requirement of Mettl3-promoted MyoD mRNA maintenance in proliferative myoblasts for skeletal muscle differentiation. *Open Biol.* (2017) 7:170119. doi: 10.1098/rsob.170119

41. Zhang H, Shi X, Huang T, Zhao X, Chen W, Gu N, et al. Dynamic landscape and evolution of m6A methylation in human. *Nucleic Acids Res.* (2020) 48:6251–64. doi: 10.1093/nar/gkaa347

42. Goulding M, Paquette A. Pax genes and neural tube defects in the mouse. *Ciba Found Symp.* (1994) 181:103–13. doi: 10.1002/9780470514559.ch7

43. Relaix F, Montarras D, Zaffran S, Gayraud-Morel B, Rocancourt D, Tajbakhsh S, et al. Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *J Cell Biol.* (2006) 172:91–102. doi: 10.1083/jcb.200508044

44. Alarcon CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. N^6-methyladenosine marks primary microRNAs for processing. *Nature.* (2015) 519:482–5. doi: 10.1038/nature14281

45. Min KW, Zealy RW, Davila S, Fomin M, Cummings JC, Makowsky D, et al. Profiling of m6A RNA modifications identified an age-associated regulation of AGO2 mRNA stability. *Aging Cell.* (2018) 17:e.12753. doi: 10.1111/acel.12753

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 04 August 2022
doi: 10.3389/fgene.2022.935238

Integrative Analysis Between Genome-Wide Association Study and Expression Quantitative Trait Loci Reveals Bovine Muscle Gene Expression Regulatory Polymorphisms Associated With Intramuscular Fat and Backfat Thickness

*Bárbara Silva-Vignato[1], Aline Silva Mello Cesar[2], Juliana Afonso[3], Gabriel Costa Monteiro Moreira[1], Mirele Daiana Poleti[4], Juliana Petrini[1], Ingrid Soares Garcia[1], Luan Gaspar Clemente[1], Gerson Barreto Mourão[1], Luciana Correia de Almeida Regitano[3] and Luiz Lehmann Coutinho[1]**

[1]Department of Animal Science, College of Agriculture "Luiz de Queiroz", University of São Paulo, Piracicaba, Brazil, [2]Department of Agroindustry, Food, and Nutrition, College of Agriculture "Luiz de Queiroz", University of São Paulo, Piracicaba, Brazil, [3]Embrapa Pecuária Sudeste, São Carlos, Brazil, [4]College of Animal Science and Food Engineering, University of São Paulo, Pirassununga, Brazil

Understanding the architecture of gene expression is fundamental to unravel the molecular mechanisms regulating complex traits in bovine, such as intramuscular fat content (IMF) and backfat thickness (BFT). These traits are economically important for the beef industry since they affect carcass and meat quality. Our main goal was to identify gene expression regulatory polymorphisms within genomic regions (QTL) associated with IMF and BFT in Nellore cattle. For that, we used RNA-Seq data from 193 Nellore steers to perform SNP calling analysis. Then, we combined the RNA-Seq SNP and a high-density SNP panel to obtain a new dataset for further genome-wide association analysis (GWAS), totaling 534,928 SNPs. GWAS was performed using the Bayes B model. Twenty-one relevant QTL were associated with our target traits. The expression quantitative trait loci (eQTL) analysis was performed using Matrix eQTL with the complete SNP dataset and 12,991 genes, revealing a total of 71,033 cis and 36,497 trans-eQTL (FDR < 0.05). Intersecting with QTL for IMF, we found 231 eQTL regulating the expression levels of 117 genes. Within those eQTL, three predicted deleterious SNPs were identified. We also identified 109 eQTL associated with BFT and affecting the expression of 54 genes. This study revealed genomic regions and regulatory SNPs associated with fat deposition in Nellore cattle. We highlight the transcription factors *FOXP4, FOXO3, ZSCAN2,* and *EBF4,*

Abbreviations: BFT, backfat thickness; ECM, extracellular matrix; eQTL, expression quantitative trait loci; IMF, intramuscular fat; LT, Longissimus thoracis; QTL, quantitative trait loci; TF, transcription factor.

OPEN ACCESS

Edited by:
Nuno Carolino,
Instituto Nacional Investigaciao Agraria e Veterinaria (INIAV), Portugal

Reviewed by:
Guillermo Giovambattista,
CONICET Institute of Veterinary Genetics (IGEVET), Argentina
José Marcelo Soriano Viana,
Universidade Federal de Viçosa, Brazil

***Correspondence:**
Luiz Lehmann Coutinho
llcoutinho@usp.br

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

Received: 03 May 2022
Accepted: 23 June 2022
Published: 04 August 2022

Citation:
Silva-Vignato B, Cesar ASM, Afonso J, Moreira GCM, Poleti MD, Petrini J, Garcia IS, Clemente LG, Mourão GB, Regitano LCdA and Coutinho LL (2022) Integrative Analysis Between Genome-Wide Association Study and Expression Quantitative Trait Loci Reveals Bovine Muscle Gene Expression Regulatory Polymorphisms Associated With Intramuscular Fat and Backfat Thickness. Front. Genet. 13:935238. doi: 10.3389/fgene.2022.935238

involved in lipid metabolism-related pathways. These results helped us to improve our knowledge about the genetic architecture behind important traits in cattle.

Keywords: backfat thickness, carcass and meat quality, expression quantitative trait loci, intramuscular fat content, RNA-Seq, SNP, Nellore cattle

INTRODUCTION

Over the last years, several studies have shown a growing interest in understanding the molecular mechanisms regulating carcass and meat quality traits in beef cattle (Cesar et al., 2014; Fernandes Júnior et al., 2016; Silva-Vignato et al., 2017; Park et al., 2018; Raza et al., 2020). Carcass and meat quality fat traits, such as backfat thickness (BFT) and intramuscular fat content (IMF) are economically important for the beef industry since they affect the yield of cuts, dressing percentage, and the final consumer perception of meat quality (Yokoo et al., 2008; Park et al., 2018). BFT and IMF are traits related to the final amount of fat in the carcass and play an important role in the determination of meat palatability. The subcutaneous fat layer protects the carcass in the cooling process, minimizing evaporative weight loss and avoiding muscle fiber cold-shortening (Yokoo et al., 2008). The IMF content, also known as marbling, affects beef juiciness, tenderness, and palatability, relevant sensory characteristics for the consumers (Troy et al., 2016; Park et al., 2018). Moreover, the IMF is composed of higher levels of polyunsaturated fatty acid (PUFA) and monounsaturated fatty acid (MUFA), which are beneficial for human health (Troy et al., 2016).

Fat deposition in beef cattle depends on several intrinsic and extrinsic factors, such as the stage of growth, physiological maturity, nutrition, and genetics (Kauffman and Berg, 2011). Researchers have already explored some of these factors showing the differences in body composition among and within breeds (Burrow et al., 2001; Yokoo et al., 2008; Lopes et al., 2012). In the genetics field, genome-wide association studies (GWAS) have been used to detect DNA variants and genomic regions (quantitative trait loci, QTL) associated with carcass and meat quality traits (Tizioto et al., 2013; Cesar et al., 2014; Fernandes Júnior et al., 2016; Raza et al., 2020). Fernandes Júnior et al. (2016), identified genomic regions and putative candidate genes associated with ribeye area and BFT in Nellore cattle. Recently, Martins et al. (2021) found genomic regions on chromosomes 1, 2, 5, 6, 7, 8, 10, 13, 14, and 26, which together explained 12.96% of the total additive genetic variance of fatness (backfat and rump fat thickness) in Nellore cattle. The authors reported seven candidate genes involved in metabolic pathways related to fatness and lipid metabolism (Martins et al., 2021). In previous studies from our research group, Tizioto et al. (2013) found a small effect QTL associated with meat and carcass quality traits in Nellore cattle. Cesar et al. (2014) identified 23 moderate effect QTL associated with fatty acids composition and small effect QTL associated with intramuscular fat in Nellore cattle. Although GWAS reveal genomic regions and putative candidate genes associated with the phenotypes, such analysis provides limited information on the molecular regulation of phenotypes (Michaelson et al., 2009).

Understanding the architecture of gene expression is fundamental to unraveling the molecular mechanisms regulating complex traits (Spielman et al., 2007; Lee, 2018). In previous studies from our lab, we have identified differentially expressed genes in the muscle transcriptome of Nellore cattle with extreme values for BFT and IMF, revealing metabolic pathways and biological processes involved with these traits (Cesar et al., 2015; Silva-Vignato et al., 2017). We also detected modules of co-expressed genes correlated with BFT in Nellore cattle, underling relevant pathways involved in bovine fat deposition (Silva-Vignato et al., 2019). These studies helped us to gain insights into how gene expression influenced these phenotypes. However, there are still gaps in our knowledge about gene expression regulation in cattle.

The expression quantitative trait loci (eQTL) mapping effectively integrates genetic variations and gene expression at the whole-genome level (Westra and Franke, 2014). eQTL data provide substantial insights into transcriptional regulation, functional interpretation for trait-associated SNP, and genetic factors that regulate a specific disease or a complex phenotype (Michaelson et al., 2009; Shabalin, 2012; Westra and Franke, 2014). Thus, our main goal was to identify gene expression regulatory polymorphisms within genomic regions associated with intramuscular fat and backfat thickness in Nellore cattle. To achieve this goal, GWAS and eQTL analyses were performed using an SNP dataset formed by transcribed variants mined from RNA-Seq data combined with a high-density panel of SNP.

MATERIALS AND METHODS

Animals, Samples, and Phenotypes

The experimental procedures related to animal handling and care were approved by the Institutional Animal Care and Use Committee Guidelines from EMBRAPA (CEUA 01/2013).

A population of 193 Nellore steers, derived from an experimental herd of the Brazilian Agricultural Research Corporation (EMBRAPA), and originated from 34 unrelated bulls representing the principal Brazilian Nellore genealogies, was used in the current study. Between the years 2009 and 2011, the animals were raised in grazing systems and finished in feedlots with the same handling and nutritional conditions. The steers were slaughtered at an average age of 25 months and 452 kg in a commercial slaughterhouse located in Bariri (São Paulo, Brazil), following the Brazilian Ministry of Agriculture, Livestock and Food Supply (MAPA) regularization. More details are provided elsewhere (Tizioto et al., 2013; Cesar et al., 2014).

For the RNA-Seq, a *Longissimus thoracis* (LT) muscle sample of approximately 5 g was collected from the right side of each carcass between the 12th and 13th ribs immediately after the

animal's death and stored in ultra-freezer at −80°C until the analysis. For measurements of intramuscular fat content (IMF, %) and backfat thickness (BFT, mm), a beef sample of the LT muscle (12th–13th ribs, left side of the carcass) was collected 24 h after slaughter. For IMF analysis, beef samples of approximately 100 g were lyophilized and ground, then IMF was achieved using the Ankom XT20 extractor, following the AOCS protocol (AOCS, 2004), a more complete description can be found in Cesar et al. (2014). The BFT was measured by using a graduated ruler, more details in Tizioto et al. (2013).

High-Density Genotyping Data

The high-density genotyping data acquisition was already described elsewhere (Cesar et al., 2014). Briefly, the genotyping analysis was performed at the Bovine Functional Genomics Laboratory ARS/United States and ESALQ Genomics Center (Piracicaba, São Paulo, Brazil), using the BovineHD 770 k BeadChip (Infinium BeadChip, Illumina, San Diego, CA, United States) following Illumina's protocol. As a quality control step, SNPs with call rate ≤ 95%, minor allele frequency (MAF) ≤ 5%, located in sexual chromosomes, and those not mapped in the *Bos taurus* ARS-UCD1.2 reference genome were excluded from further analysis.

RNA-Sequencing

For total RNA extraction, a sample of 100 mg of the LT muscle was processed using the Trizol reagent (Life Technologies, Carlsbad, CA, United States), following the manufacturer's guidelines. After extraction, RNA integrity was verified using the Bioanalyzer 2100 (Agilent, Santa Clara, CA, United States), and the samples presenting RNA integrity numbers (RIN) greater than 7 were considered for the next analyses. A total of 2 μg of RNA from each sample was used for the cDNA library preparation, according to the protocol described in the TruSeq RNA Sample Preparation kit v2 guide (Illumina, San Diego, CA, United States). The libraries were sequenced using the Illumina HiSeq2500 ultra-high-throughput sequencing system with the TruSeq SBS kit v3-HS (200 cycles), as described in Cesar et al. (2015). All sequencing analyzes were performed at ESALQ Genomics Center (Piracicaba, São Paulo, Brazil). After sequencing, the SeqyClean package v. 1.4.13 (Zhbannikov et al., 2017) was utilized to remove low-complexity reads and the adapters sequences from the library preparation step. For the quality control visualization, FastQC software v. 0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used.

The read alignment against the bovine reference genome *Bos taurus* ARS-UCD1.2 was carried out using STAR (Spliced Transcripts Alignment to a Reference) (Dobin and Gingeras, 2015) v. 2.7 with Ensembl (release 96) gene annotation file. To count the reads, we applied the HTSeq software (Anders et al., 2015) v. 0.11.1 inside STAR. Only reads that were exclusively mapped to known chromosomes were used in this study.

Read counts for each gene were normalized to CPM (Counts per million) using the Bioconductor package edgeR (Robinson et al., 2010), and then the CPM values were log2 transformed (log2-CPM). Genes expressed at a low level or not expressed (log2-CPM value < 0) and expressed in <50% of the samples, were filtered out from the analysis. Additionally, concerning technical biases affecting gene expression, a batch effect correction was performed using the NOISeq R Package (Tarazona et al., 2015) v. 2.16.0. For that, a Principal Component Analysis (PCA) allowed us to explore the dataset and detect possible batch effects. Then, using the ARSyNseq function, we filtered out the noise associated with the batch effect, a combination of flow cell and lane. The datasets analyzed in this study can be found in the European Nucleotide Archive (ENA) repository (EMBL-EBI) under the accession codes: PRJEB13188, PRJEB10898, and PRJEB19421.

Variant Calling Analysis and SNP Annotation

For the variant calling analysis in the muscle transcriptome, the Genome Analysis Toolkit (GATK) v. 4.1.0.0 was used in the Genomic Variant Call Format (GVCF) mode (Brouard et al., 2019). Using this approach, all genotypes' types were reported in a final VCF file. The variants were called following the GATK Best Practices, and the Ensembl *Bos taurus* dbSNP (release 96) was used as known variants. The *HaplotypeCaller* algorithm was used to call the variants individually, generating GVCF files for each sample. These files were then merged using the *CombineGVCF* tool, and the joint genotyping analysis was performed using the *GenotypeGVCF*. In the end, a VCF file with all samples genotyped was achieved. After the variant calling, we filtered the SNP for variant quality score (QUAL) ≥ 30 and total depth of coverage (DP) > 10, using BCFtools (Li, 2011) v. 1.9. Moreover, the SNP with call rate < 95%, MAF < 5%, located in sexual chromosomes, and non-biallelic, were removed from the SNP dataset. The variants' annotation and functional consequences were predicted using the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016) v. 95.2.

Genome-Wide Association Study

Previously to the association study, the filtered SNPs from the RNA-Seq variant calling and the Bovine HD BeadChip were combined into one complete dataset. BEDTools (Quinlan and Hall, 2010) v. 2.27.1 was used to check for common variants (located in the same genome position) in the two datasets. Then, the transcribed variants located in the same spot as those from the Bovine HD BeadChip were removed from the analysis. Thus, a complete dataset containing all SNPs (unique variants) was used for the following analysis.

The GWAS was performed using the GenSel software (Garrick and Fernando, 2013) with a Bayesian approach. First, a Bayes C model was used to estimate the prior genetic and residual variances for each trait with a calculated π (0.9997). Then, these values were used as priors to run a Bayes B model, as previously described in (Cesar et al., 2014; Moreira et al., 2018). The mathematical model was

$$\boldsymbol{y} = \boldsymbol{Xb} + \sum_{j=1}^{k} \boldsymbol{a}_j \beta_j \delta_j + \boldsymbol{e},$$

where $\boldsymbol{y}$ was the vector of phenotypic values, $\boldsymbol{X}$ represented the incidence matrix for fixed effects, $\boldsymbol{b}$ was the vector of fixed effects, K was the number of SNP variants (534,928), $\boldsymbol{a_j}$ was the column vector representing the SNP covariate at locus j, assumed to be normally distributed $N(0, \sigma^2_\beta)$ when $\delta_j = 1$, but $\beta_j = 0$ when $\delta_j = 0$, with δ_j being a random variable 0/1, indicating the absence

(probability π) or presence (probability 1-π) of *locus j* in the model, and **e** represented the vector of residuals associated with the analysis. In the model, the contemporary group (animals from the same farm, year, and slaughter date) was set as a fixed effect and hot carcass weight as a covariate.

The GenSel program uses Markov-Chain Monte Carlo (MCMC) to estimate the effect of each SNP among all SNPs in each interaction. In this procedure, 41,000 interactions, with the first 1,000 interactions being discarded, were accumulated to obtain the posterior mean effect for each SNP. A map file was used to position the SNPs into 2,502 non-overlapping 1 Mb SNP windows. Manhattan plots containing the variance explained by each SNP window along the autosomal chromosomes were constructed for each trait. Based on an infinitesimal model (Onteru et al., 2013; Van Goor et al., 2016), it is expected that each window explains 0.04% (100%/2,502) of the genetic variance, thus, windows explaining five times more than the expected (0.20%) were considered as relevant QTL regions.

The Cattle QTL database (Hu et al., 2019) (Cattle QTLdb, release 41) was used to search for known QTL that could be overlapping our relevant QTL regions. For that, we used an in-house R script and the BED file with QTL coordinates according to the *Bos taurus* ARS-UCD1.2 genome, available on Cattle QTLdb. We also checked for previously detected QTL reported by our research group for the interest traits (Tizioto et al., 2013; Cesar et al., 2014). Before this, the LiftOver tool from UCSC Genome Browser (Casper et al., 2018) was used to convert the genome coordinates from the *Bos taurus* UMD3.1 for the current version ARS-UCD1.2. Finally, the genes within the relevant QTL were annotated using the Ensembl Biomart (Ensembl Genes 100).

Expression Quantitative Trait Loci Identification and Functional Annotation

The R package Matrix eQTL (Shabalin, 2012) v. 2.3 was used to perform cis and trans-eQTL identification, using the complete dataset of SNP and genes with expression values in log2-CPM. The contemporary group was considered in the model for confounding effect correction. According to previous work from our group (Mudadu et al., 2016), no evidence of population stratification was verified in this Nellore population. In the present study, cis-eQTL were defined as SNP located no more than 1 Mb upstream or downstream from the regulated gene, and trans-eQTL as the SNP located more than 1 Mb from the regulated gene. Matrix eQTL tests for associations between SNP genotypes and gene expression using linear regression to associate each gene-SNP pair, considering additive genotype effects. The program also calculates the false discovery rate (FDR), based on Benjamini–Hochberg methodology (Benjamini and Hochberg, 1995), separately for cis and trans-eQTL (Shabalin, 2012). The lists of cis and trans-eQTL (FDR < 0.05) were annotated separately, by using VEP (McLaren et al., 2016) v. 95.2. At last, to verify if the eQTL can be affecting transcription factor (TF), we compared our results with the manually curated list of bovine TFs published by our research group (de Souza et al., 2018).

Overlap Between Relevant Quantitative Trait Loci and the Expression Quantitative Trait Loci

To verify if the SNPs within relevant QTL regions were also affecting the gene expression, an overlap analysis using those SNPs and the list of cis and trans-eQTL (FDR < 0.05) were performed utilizing the GNU/LINUX environment. Considering all the eQTL (cis and trans) within QTL regions, we used PLINK (Purcell et al., 2007) v. 1.9 to perform linkage disequilibrium (LD) pruning. The parameters applied to variant pruning were pairwise linkage with a minimum r^2 of 0.5 and window size of 100 SNPs, shifting 10 SNPs at each step. Then, carrying out the most representative eQTL (tag-SNP) and the genes regulated by them, we used Cytoscape software (Shannon et al., 2003) to build SNP-gene regulation networks for each trait. Finally, to find the molecular pathways in which the genes regulated and containing the representative eQTL were involved, we used MetaCore software (https://portal.genego.com/) from Clarivate (London, GBR) with the *Mus musculus* database.

RESULTS

Phenotypes, High-Density SNP Data, and RNA-Seq Data

In the current study, the Nellore steer population presented mean phenotypic values of 2.93% for IMF and 6.86 mm for BFT, with genomic heritability of 0.25 and 0.18 for IMF and BFT, respectively. From the high-density genotyping SNP data, we obtained a total of 414,879 SNPs that passed all the quality control filters (MAF ≥ 5%; call rate ≥ 95%; not in sex chromosomes; and mapped to the reference genome). Concerning the RNA-Seq data, an average of 18.45 million reads per animal were used as input for mapping to the *Bos taurus* ARS-UCD1.2 genome. In this analysis, 85.71% of the reads were uniquely mapped to the reference genome. **Supplementary Table S1** shows all the mapping reads statistics. Read counts were normalized to counts per million (CPM) and log2 transformed. Filtering steps were applied to remove low-level or not expressed genes and a batch effect correction was performed to minimize technical biases affecting gene expression, totaling 12,991 genes with log2-CPM expression values used in the eQTL identification.

RNA-Seq Variant Calling Analysis Revealed 120,049 Unique Variants.

The GATK variant calling analysis allowed us to identify 123,300 SNPs that passed all the quality control filters (MAF >5%; call rate > 95%; not in sex chromosomes; and biallelic). There were 3,251 SNPs that overlapped the SNP panel used for genotype (Illumina HD BeadChip). These were used to validate our SNP discovery and revealed 95.85% concordance between genotypes. Variants in the same position as the high-density SNP panel were removed for the next analysis due to their high genotype similarity, and we proceeded with 120,049 unique variants.

Functional annotation analysis of the 120,049 SNPs revealed 9,018 novel variants. **Supplementary Figure S1** shows the variant

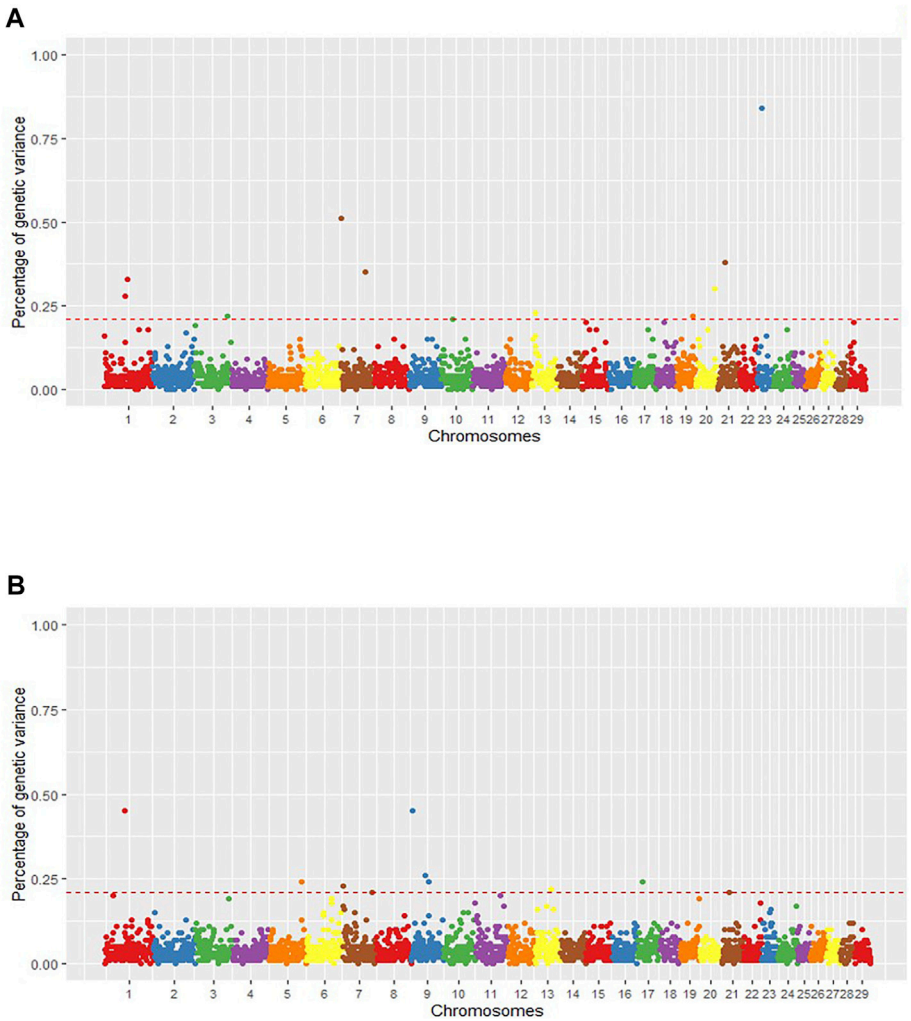


FIGURE 1 | Manhattan plot of the posterior means of the percentage of genetic variance explained by each 1 Mb SNP window across the 29 autosomal chromosomes for intramuscular fat content (IMF) **(A)** and backfat thickness (BFT) **(B)**. The X-axis represents the chromosomes, and the Y-axis, the percentage of genetic variance explained by each SNP window. Red dashed lines delimit the relevant QTL regions.

distribution across the 29 *Bos taurus* autosomal chromosomes (BTA) and the most severe consequences predicted by VEP. Most SNPs were in 3′UTR regions (23.24%), followed by downstream genes (18.82%) and intron variants (12.67%). Besides these, 25.28% were synonymous variants, while 9.81% were classified as missense mutations. The SIFT score predicted 2.85% of deleterious SNPs. A complete overview of the results obtained in the annotation analysis can be seen in **Supplementary Table S2**.

Twenty-One Relevant Quantitative Trait Loci Associated With Intramuscular Fat and Backfat Thickness

The Illumina bovine HD SNP panels were designed to cover the entire genomes with equally spaced polymorphic SNPs across different breeds. In the current study, 414,879 SNPs from the *Bos taurus* genome passed all the quality control filters. To empower the SNP database with coding and regulatory variants, we combined the Illumina panel with transcribed variants obtained from the RNA-Seq calling analysis (120,049 SNPs). As a result, for the GWAS, a total of 534,928 SNPs were associated with the phenotypes. **Figure 1** shows the Manhattan plots of the proportion of genetic variance explained by the 1 Mb SNP windows for each trait (a complete overview of the SNP windows can be seen in **Supplementary Table S3**).

Starting with IMF, we found eleven relevant QTL windows positioned on BTA1, 3, 7, 10, 13, 19, 20, 21, and 23 (**Table 1**). The QTL that explained the highest proportion of the genetic variance of the phenotype (Vg) for IMF (0.84%) was located on BTA23 at 15 Mb. **Supplementary Table S4** presents the individual SNP effects within each of the relevant windows. Additionally, using the Ensembl Genes database, we annotated the 155 genes within the relevant QTL associated with IMF (**Supplementary Table S5**). Then, we consulted the Cattle QTLdb to identify relevant QTL overlapping with known cattle QTL. Our relevant QTL

TABLE 1 | Characterization of the relevant QTL regions associated with intramuscular fat content (IMF) and backfat thickness (BFT) in a Nellore cattle population.

Traits	Chr_Mb	First—last position	Proportion Vg (%)	N SNP/window	N Genes/window
IMF	23_15	15000692–15992208	0.84	185	27
	7_2	2003622–2999865	0.51	941	26
	21_22	22001027–22990406	0.38	223	18
	7_81	81002985–81999643	0.35	284	7
	1_75	75000671–75968766	0.33	257	3
	20_62	62002061–62998209	0.30	303	11
	1_67	67000099–67991943	0.28	251	12
	13_13	13000910–13992863	0.23	208	2
	19_54	54022589–54994606	0.22	309	12
	3_108	108002916–108996760	0.22	258	16
	10_37	37022308–37997882	0.21	427	21
BFT	9_4	4040113–4999917	0.45	196	2
	1_63	63008675–63998464	0.45	201	4
	9_46	46000672–46993133	0.26	184	2
	5_104	104002715–104990551	0.24	302	8
	17_18	18002778–18997485	0.24	335	13
	9_55	55005021–55998978	0.24	144	3
	7_3	3000052–3998084	0.23	351	24
	13_52	52001168–52998893	0.22	212	32
	7_96	96002065–96995101	0.21	523	9
	21_23	23014366–23998400	0.21	224	9

Chr_Mb = map position (chromosome and position in Mb) based on the Bos taurus ARS-UCD1.2; Proportion Vg (%) = Proportion of genetic variance explained by 1 Mb SNP, window; N SNP/ window = number of SNP within the genomic region; N Genes/ window = number of genes annotated within the (Ensembl Genes 100) SNP window.

regions were previously associated with carcass and meat quality traits, such as marbling score, fatty acid composition, shear force, and body weight gain in taurine and zebuine breeds (**Supplementary Table S6**).

For BFT, we found 10 relevant QTL (**Table 1**) with 106 genes annotated on these regions (**Supplementary Table S5**). The QTL that explained the highest proportion of Vg were located on BTA9 at 4 Mb and BTA1 at 63 Mb (0.45% of Vg each). **Supplementary Table S7** reports the individual SNP effects within the relevant QTL associated with BFT. Comparing our SNP windows with known bovine QTL (QTLdb) revealed genomic regions associated with body composition and carcass quality traits, like body weight, subcutaneous fat, and shear force in taurine and zebuine cattle breeds (**Supplementary Table S6**). Moreover, looking for our group's previously identified QTL, we found two IMF QTL (BTA1_75 and BTA20_62) and three BFT QTL (BTA1_63, BTA7_3, and BTA7_96) encompassing genomic regions associated before with ribeye area, BFT, and fatty acids content (Eicosadienoic acid and Docosahexaenoic acid) in this Nellore population (Tizioto et al., 2013; Cesar et al., 2014).

Most Cis and Trans-Expression Quantitative Trait Loci Were in Transcribed Regions

We identified 71,033 cis-eQTL and 36,497 trans-eQTL (FDR < 0.05) distributed along the genome, with 5,718 SNPs acting both as cis and trans-eQTL. **Figure 2** illustrates the eQTL distribution and gene positions (Mb) along the 29 BTAs. Regarding gene regulation, 4,871 genes had their expression affected by cis-eQTL, and among them, 128 were in the curate list of bovine TFs (de Souza et al., 2018). Moreover, 6,370 genes were affected by trans-eQTL, and within them, 259 were TFs. From the total of genes, 2,560 were affected by both cis and trans-eQTL. **Supplementary Table S8** displays the complete list of cis and trans-eQTL (FDR < 0.05) and the genes regulated by them, highlighting the TFs.

VEP analysis showed that most of the cis and trans-eQTL were located on BTA19 and BTA23, respectively, while BTA20 had fewer local and distant variants (**Supplementary Figure S2**). Moreover, most eQTL were variants called from the RNA-Seq data (approximately 92% of the cis and 70% of the trans-eQTL), so the functional annotation results are similar to those presented for the complete dataset of transcript SNPs. Both cis and trans-eQTL were predominantly located in 3'UTR, intronic, and downstream gene regions. Among them, 23.86% and 22.25% were predicted to be synonymous variants, whereas 9.62% and 8.70% were classified as missense for the cis and trans-eQTL, respectively (**Supplementary Figure S3**).

Regulatory Polymorphisms Associated With Intramuscular Fat and Backfat Thickness

To identify eQTL that could be associated with our phenotypes, we overlapped the eQTL and GWAS results. Our analysis revealed that 231 and 109 eQTL variants were located on relevant QTL associated with IMF and BFT, respectively. Within the 231 eQTL associated with IMF (relevant QTL windows described in **Table 1**), 156 were cis, 26 trans, and 49 cis and trans variants. These regulatory polymorphisms affected the expression of 117 genes, including seven TFs: *ARNT, FOXO3, FOXP4, NFYA, ZFP2, ZNF354C*, and

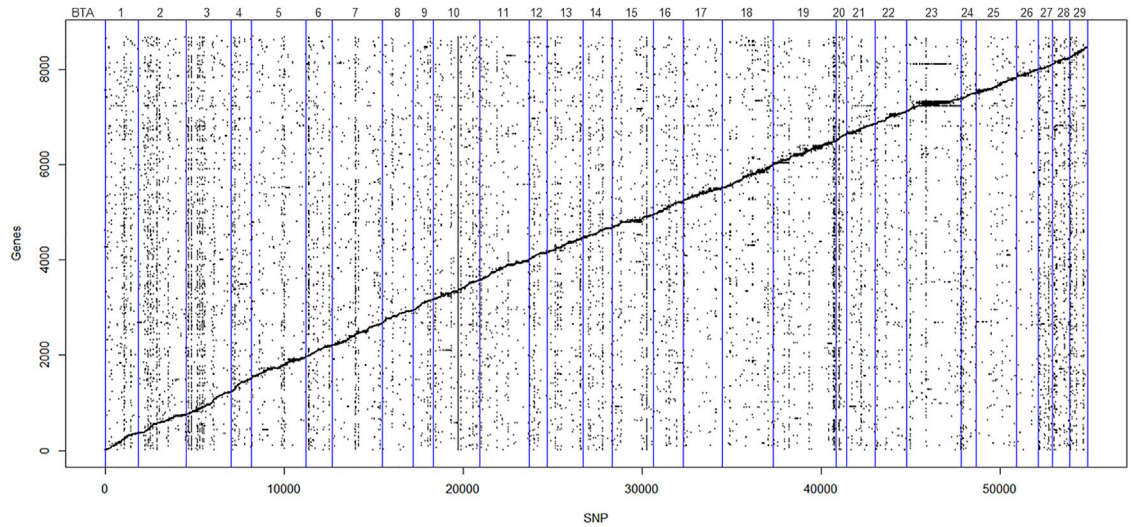


FIGURE 2 | Scatter plot of the affected genes and eQTL (FDR < 0.05). The *Y*-axis represents gene order in relation to chromosome position in the *Bos taurus* genome, and *X*-axis represents the SNP order in relation to chromosome position in the *Bos taurus* genome. Points scattered diagonally indicates cis-eQTL. Points scattered vertically indicate trans-eQTL. The vertical blue lines denote individual autosomal chromosomes.

ZSCAN2. Besides that, 12 eQTL were missense mutations (**Supplementary Table S2**), spanning the QTL regions located on BTA1, 3, 7, 10, 19, and 21. Among them, the cis-eQTL rs381713284 (BTA21 at 22 Mb, SIFT = 0.05), the trans-eQTL rs379524684 (BTA3 at 108 Mb, SIFT = 0.03), and the trans-eQTL rs110129172 (BTA10 at 37 Mb, SIFT = 0.02) were deleterious SNP. The SNP rs382320484 (BTA10 at 37 Mb), upstream of the *TMEM87A* and *GANC* genes, affected the expression of the largest number of genes (12 genes in trans and two genes in cis); followed by the novel SNP 21:22425675 (chromosome: position), a synonymous variant located on the *ZNF592* gene, that affected the expression of nine genes (seven genes in trans and two genes in cis), being two TFs (*FOXO3* and *ARNT*). Moreover, within the genes regulated by eQTL associated with IMF, the two genes regulated by the larger number of eQTL were the pseudogene *ENSBTAG00000052719* (BTA7) by 21 cis-eQTL, and the *TMEM87A* (BTA10) by 13 cis-eQTL. In **Figure 3**, we presented the SNP-gene regulation networks for the eQTL associated with IMF, focusing on the eQTL affecting TFs and their direct connections.

Associated with BFT, there were 74 cis, 20 trans, and 15 both cis and trans-eQTL. These variants were located on the SNP windows BTA9_46, BTA5_104, BTA17_18, BTA7_3, BTA13_52, BTA7_96, and BTA21_23 (see **Table 1**). Together, the 109 eQTL affected the expression levels of 54 genes, among them two TFs, the *EBF4* and *ZSCAN2*. The genes regulated by the larger number of eQTL were *NDUFC1* (BTA17), which is affected by nine eQTL (cis and trans), and the novel gene *ENSBTAG00000025383* (BTA12), which is affected by eight eQTL (cis and trans). **Figure 4** illustrates the SNP-gene regulation networks for the eQTL associated with BFT, focusing on the eQTL affecting TFs and their direct connections. As occurred for IMF, we found 12 missense variants within our relevant eQTL. These variants were located on BTA5, 7, 13, 17, and 21 and all of them were classified as tolerated by VEP (without deleterious effects). The complete list of the eQTL spanning IMF and BFT relevant QTL, their regulated genes, and beta-values (effect size and direction) are presented in **Supplementary Table S9**.

At last, to investigate if the genes regulated and containing eQTL associated with the interest traits were involved in lipid metabolism-related pathways, we did an enrichment analysis of these genes (135 and 74 genes for IMF and BFT, respectively). This analysis revealed some interesting pathways, such as signal transduction, cell cycle, development, and transport, underlining the AKT and WNT signaling. **Figure 5** shows the top ten Pathway Maps [−log (p-value)] enriched for the genes related to IMF and BFT.

DISCUSSION

In the present study, we used an SNP dataset constituted by RNA-Seq variants and a high-density genotyping panel to perform an integrative analysis between GWAS and eQTL. The idea was to find gene expression regulatory polymorphisms associated with intramuscular fat and backfat thickness in bovine. Backfat thickness and intramuscular fat deposition are of economic importance to the beef cattle industry. The BFT is the best predictor of overall fatness in the animal's body, impacting carcass cutability and meat yield (Yokoo et al., 2008; Lopes et al., 2012). The IMF is positively correlated with beef tenderness, a meat quality trait that strongly affects consumer satisfaction and repurchase decision (Park et al., 2018). Moreover, the beef fatty acids composition is associated with human health (Troy et al., 2016). Herein, the mean phenotypic values of IMF and BFT were higher than those presented in the literature for Nellore cattle, with mean values ranging from 2 to 5 mm for BFT, and around 1% for IMF (Yokoo et al., 2010; Borges et al., 2014;

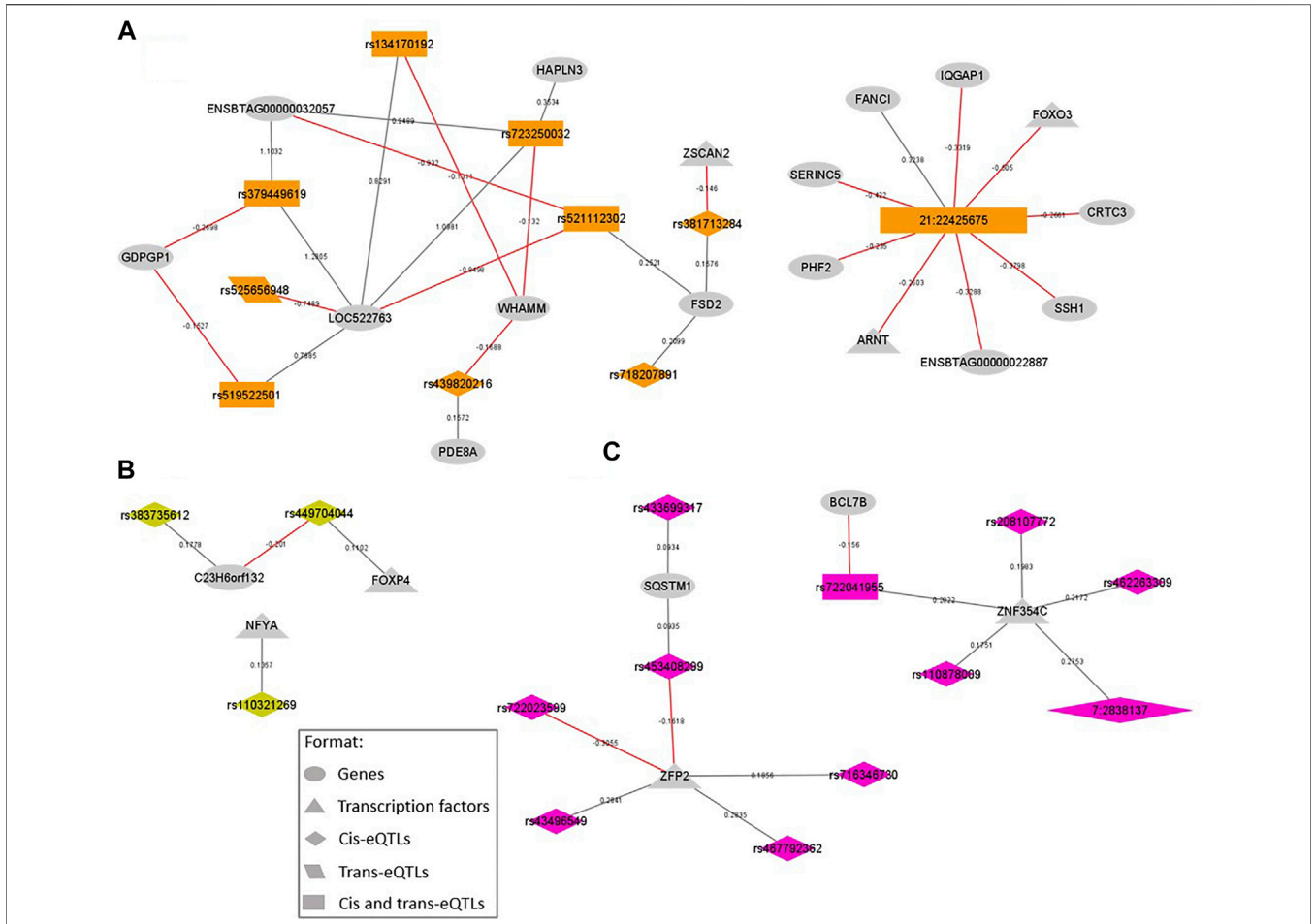


FIGURE 3 | SNP-gene regulation networks representing the eQTL variants located within QTL windows associated with intramuscular fat content (IMF) and the genes regulated by them, focusing on the variants regulating transcription factors (TFs) and their direct connections. The colors are coded by QTL window: **(A)** BTA21_22, eQTL represented in orange; **(B)** BTA23_15, eQTL represented in lightgreen; **(C)** BTA7_2, eQTL represented in pink. All the formats are described in the legend. New variants are represented by chromosome: position. Gray lines represent a positive beta-value and red ones represent a negative beta-value.

Fernandes Júnior et al., 2016). Nevertheless, selection for these traits can be difficult as they are expressed later in the animal's life. In this way, identifying genes and genetic markers with causative effects over these traits will help improve breeding progress in bovine (Fernandes Júnior et al., 2016).

GWAS has been widely used to identify genetic variants and putative candidate genes associated with complex phenotypes (Cesar et al., 2014; Fernandes Júnior et al., 2016; Raza et al., 2020; Martins et al., 2021). However, panels used for GWAS studies are not designed to have causative SNP since the goal is to have informative markers across the genome (Tam et al., 2019). In the current study, we used RNA-Seq-based SNP to empower the high-density genotyping panel. According to Suárez-Vega et al. (2015), calling variants from RNA-Seq raises the chances of discovering causative mutations harboring or neighboring QTL and can provide a better understanding of the regulatory mechanisms underlying eQTL.

With the incorporation of RNA-Seq-based SNP, we were able to identify QTL associated with IMF and BFT that better explained the genetic variance of the phenotype compared with previous works (Tizioto et al., 2013; Cesar et al., 2014), even using a minor subset of the population (193 animals that have RNA-Seq information). Tizioto et al. (2013), using only the high-density SNP panel (Bovine HD 770 k) and 536 animals from the same population, found that the highest effect QTL for BFT was located on BTA11 and only explained 0.36% of Vg, with genomic heritabily (h^2) of 0.21. Here, the two QTL with the highest effect for BFT were detected on BTA9 and BTA1, and each explained 0.45% of Vg, with a slightly smaller h^2 of 0.18. In another study using the same population (Cesar et al., 2014) with 386 Nellore steers with phenotypes for IMF and the HD 770 k chip, the highest effect QTL for IMF (located on BTA10) explained 0.66% of Vg (h^2 = 0.25), while in this study, the highest effect QTL was detected on BTA23 and it explained 0.84% of Vg (h^2 = 0.25).

Complementary to the GWAS analysis, we identified 71,033 cis-eQTL and 36,497 trans-eQTL in this Nellore population, most of them from the RNA-Seq dataset, showing

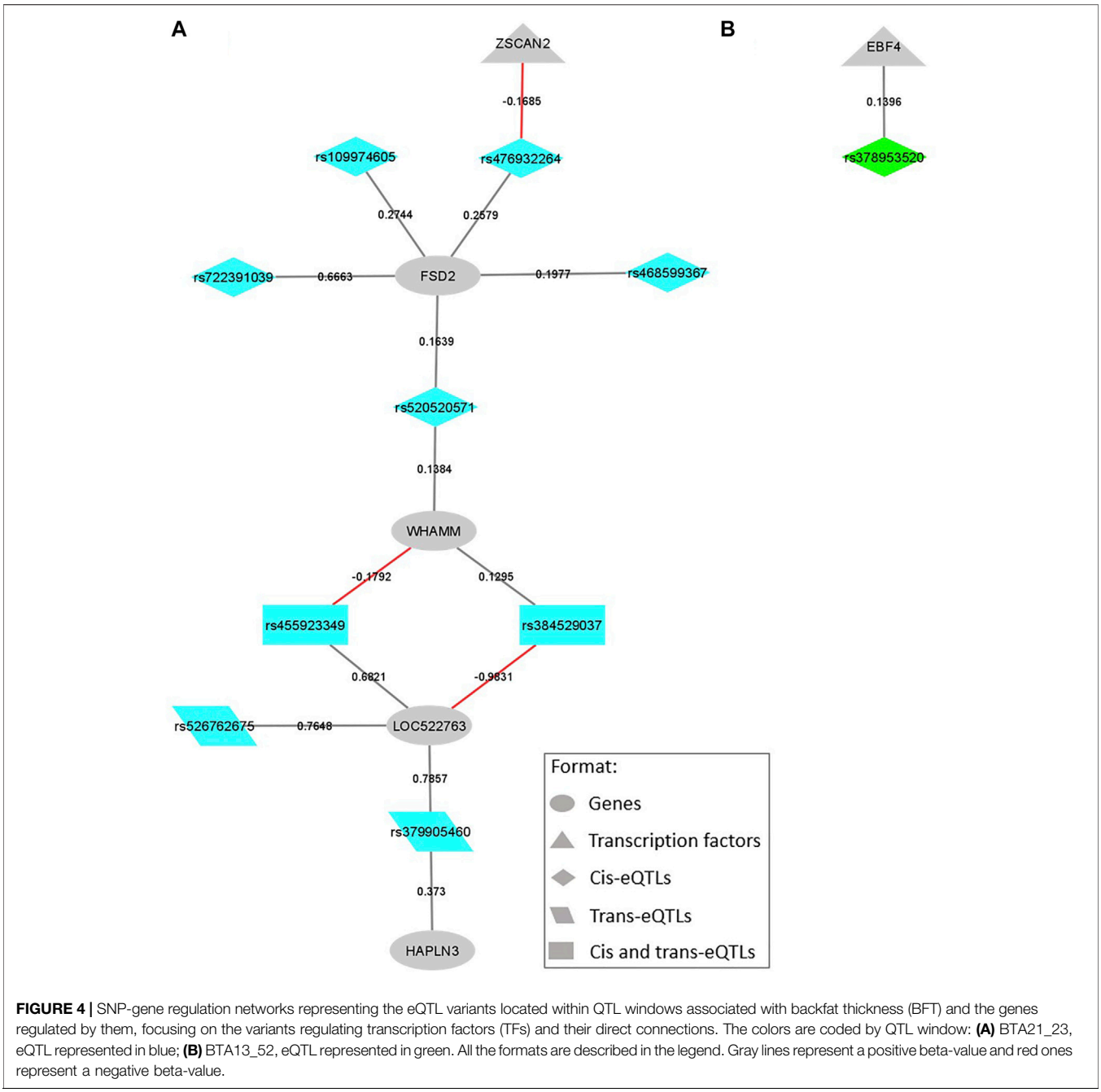


FIGURE 4 | SNP-gene regulation networks representing the eQTL variants located within QTL windows associated with backfat thickness (BFT) and the genes regulated by them, focusing on the variants regulating transcription factors (TFs) and their direct connections. The colors are coded by QTL window: **(A)** BTA21_23, eQTL represented in blue; **(B)** BTA13_52, eQTL represented in green. All the formats are described in the legend. Gray lines represent a positive beta-value and red ones represent a negative beta-value.

the importance of adding transcribed variant calling information on eQTL mapping, as stated by Suárez-Vega et al. (2015). Most of the SNPs were in 3′UTR, introns, and downstream gene regions, but some were missense variants (about 9%). It is important to keep in mind that the polymorphisms identified in this study may not be the causative mutation but could be in linkage disequilibrium with the causative one (Mueller, 2004), thus explaining how a missense SNP could be associated with gene expression variation.

There are several advantages in integrating GWAS and eQTL data. Expression quantitative trait loci are essential for understanding the genetic basis of cellular processes and complex traits (Lee, 2018). Moreover, eQTL are essential for the functional interpretation of trait-associated polymorphisms and identification of genes with expression levels associated with complex phenotypes (Westra and Franke, 2014; Littlejohn et al., 2016; Cai et al., 2019). In the current work, we focused on the overlap between the eQTL and relevant QTL regions for each

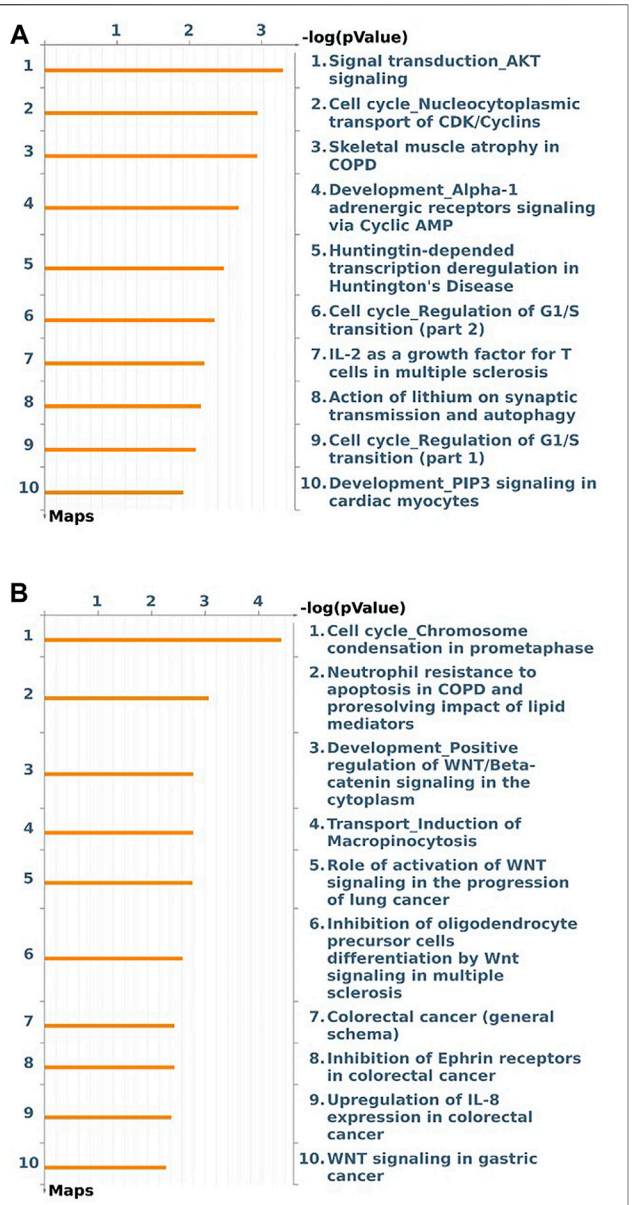


FIGURE 5 | Top 10 pathway maps enriched for the genes regulated and containing the eQTL encompassing relevant QTL regions associated with intramuscular fat content (IMF) **(A)** and backfat thickness (BFT) **(B)** in a Nellore cattle population.

trait, concentrating on the mutations affecting the expression of transcription factors.

Intramuscular Fat Content

Seven TFs were regulated by eQTL harbored on QTL regions associated with IMF. Transcription factors are cellular components that exert an essential role in regulating gene expression (Vaquerizas et al., 2009). Studying how these components are regulated is attractive to a more in-depth investigation of gene expression patterns during development and terminally differentiated cells (Calkhoven and Ab, 1996). Thus, we first highlight here the *FOXP4* TF, positively regulated by a cis-eQTL (rs449704044) located downstream of the novel gene *ENSBTAG00000054479*, and harbored on the most relevant QTL for IMF (BTA23 at 15 Mb), that explained 0.84% of Vg. The Forkhead box (Fox) family genes are expressed in various tissues, acting both in developmental processes as in tissue maintenance during adult life (Golson and Kaestner, 2016; Zhu, 2016). The same cis-eQTL (rs449704044) negatively affected the expression levels of the *C23H6orf132* gene (**Figure 3B**). Although there is a lack of information about the *C23H6orf132* gene function in cattle, according to the GeneCards human database (Stelzer et al., 2016), Fox TFs were predicted to bind TFBS (transcription factor-binding sites) of this gene. The *C23H6orf132* was also regulated by the rs383735612 cis-eQTL, located in the 3'UTR region of the *TAF8* gene. The *TAF8* is an important modulator of early adipogenesis. Because of its histone fold domain, *TAF8* can inhibit adipogenesis by specifically downregulating the expression of the peroxisome proliferator-activated receptor γ (PPARγ) and the CCAAT enhancer-binding protein α (C/EBPα), major promoters of adipogenesis (Guermah et al., 2003; Lefterova et al., 2014). Wang et al. (2013) found *TAF8* as a target gene of a miRNA highly expressed in subcutaneous fat of beef cattle, and further, enriched for the "regulation of fat cell differentiation" biological process, corroborating its relevance in controlling fat deposition in bovine. Another TF affected by an eQTL located on BTA23 was *NFYA* (see **Figure 3B**), also known as CCAAT-box Binding Protein A (CBP-A). This TF plays a role in the early development of adipocytes, as well as, is essential for leptin gene expression (Lu et al., 2015). The knockout of this gene in mice resulted in lipodystrophy with a progressive loss of adipose tissue (Lu et al., 2015).

We also emphasize another member of the Forkhead box family, the *FOXO3.* This TF is negatively regulated in trans by a synonymous novel SNP (21:22425675) located on the exonic region of *ZNF592* gene. This SNP also regulates six more genes in trans and two genes in cis (**Figure 3A**). In the skeletal muscle, FoxO genes, including *FOXO3*, are responsible for switching carbohydrate to lipid as an energy source during starvation periods and can interact with the PPARγ (Gross et al., 2008). In previous work, Cesar et al. (2016), identified *FOXO1* and *FOXO3* as upstream regulators of gene expression in the skeletal muscle of Nellore cattle influenced by a variation in oleic acid content. Gui and Jia (2018) found a polymorphism in the 3'UTR region of *FOXO1* associated with IMF in Qinchuan cattle. The authors hypothesized that the variant could affect *FOXO1* expression levels through miRNA activity, thus modulating changes in fatty acid metabolism. This hypothesis corroborates our findings of a regulatory SNP affecting the expression levels of a FoxO gene associated with IMF in beef cattle.

From the list of nine genes regulated by the aforementioned novel SNP 21:22425675 (**Figure 3A**), there is one more TF, the *ARNT*, and some noteworthy genes, such as the *CRTC3, IQGAP1,* and *SSH1*. The *ARNT*, negatively regulated in trans, is a nuclear translocator that binds the aryl-hydrocarbon receptor (AhR). This binding forms a heterodimer that attaches to Xenobiotic/Dioxin response element sequences (XRE/DRE) of different target genes, activating mRNA transcription (Ishihara et al., 2018). Conversely,

the *CRTC3* and the *IQGAP1* were regulated in cis by this eQTL. The first one, *CRTC3* is part of the cAMP responsive element-binding protein (CREB)-regulated transcription coactivator (CRTC) family and plays an important role in lipid and glucose metabolism (Liu et al., 2020; Liu et al., 2021). Studying IMF deposition, Liu et al. (2020) overexpressed *CRTC3* in porcine IMF adipocytes and observed a faster accumulation of lipid droplets in cells together with an upregulation of important fat metabolism genes, such as perilipin, PPARγ, C/EBPα, leptin, and FABP4 (Fatty acid-binding protein 4). In a more recent work, Liu et al. (2021), demonstrated that the overexpression of *CRTC3* changes the metabolic profile in intramuscular adipocytes, and also promotes adipogenic differentiation of intramuscular and subcutaneous adipocytes through the calcium signaling pathway. We also highlight that in the same genomic region (BTA21, 22 Mb), we found two other eQTL (rs379449619 and rs525656948) located on the *CRTC3* gene associated with IMF in our population, emphasizing its relevance for IMF molecular regulation. As for *IQGAP1,* this gene encodes a ubiquitously expressed scaffolding protein implicated in several cellular processes, including mitogen-activated protein kinase and AKT signaling cascades (Erickson and Anakk, 2018; Hedman et al., 2021). Studies investigating the loss of this protein in mice demonstrated a reduced PPARγ activity, as well as, a defective transcription of gluconeogenesis and fatty acid synthesis genes (Erickson and Anakk, 2018; Hedman et al., 2021). Finally, the *SSH1* gene, known by its function in cytoskeleton organization and cell migration, was previously found associated with body fat in humans (Gomez-Santos et al., 2011). In earlier work from our group, studying the proteomic profile of high and low IMF Nellore cattle, we found SSH1 protein downregulated in the group with higher values of IMF, suggesting its involvement in the cellular rearrangement needed for adipocyte growth (Poleti et al., 2018).

Still focusing on the genomic window of BTA21, we emphasize the zinc-finger gene *ZSCAN2*, a TF negatively affected in cis by a deleterious SNP (rs381713284). Deleterious mutations can be defined as genetic alterations that raise individual susceptibility or predisposition to diseases/disorders. These variations often occur in coding regions and are typically missense, causing changes in the amino acid sequence, and consequently, in the protein (Plekhanova et al., 2018). According to van Strien (2018), it is coherent that variants causing changes in the protein product (missense) also can affect the expression levels of the gene coding for the protein (cis-effects) or on other genes (trans-effects). Here the rs381713284 (located on the *WDR73* gene), which changes an Arginine for a Cysteine in the protein sequence, was the only missense eQTL presenting cis-effects. Leal-Gutiérrez et al. (2020) found that a higher expression level of *WDR73* is associated with a lower meat quality index (lower marbling score, higher connective tissue content, tougher and dryer meat) in Angus-Brahman steers. Although missense mutations on the *WDR73* gene were already reported in humans associated with neuro disorders (Jiang et al., 2017), as far as we know, in cattle, there were no previous reports of deleterious SNPs on this gene associated with IMF.

Other TF identified herein affected by eQTL associated with IMF (**Figure 3C**) were the *ZFP2* and the *ZNF354C*, both members of the zinc-finger family. Together, *ZFP2* and *ZNF354C* were affected by 10 cis-eQTL located on BTA7 at 2 Mb, the second most relevant QTL for IMF, explaining 0.51% of Vg. Zinc-finger proteins are a large family of TFs characterized by a zinc-finger domain in their structure. They are ubiquitously expressed in eukaryotic genomes, participating in growth regulation, cell development, immunity, and signal transduction pathways. During adipogenesis, zinc-finger TFs are key molecules in preadipocytes differentiation and adipocyte determination. Moreover, zinc-finger TFs can both activate and inhibit the PPARγ and C/EBPs (Wei et al., 2013; Cassandri et al., 2017). Most of the eQTL affecting these two zinc-finger gene expression levels were located in novel genes, except for two of them (rs208107772 and rs462263309) located on the exonic region of *ADAMTS2*. The A Disintegrin and Metalloproteinase with Thrombospondin motifs (ADAMTS) family exerts a principal role in the extracellular matrix (ECM) maintenance and remodeling, mainly by participating in collagen biosynthesis (Kelwick et al., 2015). Lee et al. (2010) found the *ADAMTS4* overexpressed in the *Longissimus dorsi* muscle of Korean cattle presenting high IMF. Cao et al. (2017) identified the *ADAMTS2* gene as differentially methylated and differentially expressed when comparing two sheep breeds known by their different carcass weight and meat yield, confirming that this gene may indirectly affect marbling through collagen synthesis.

Finally, the enrichment analysis of the genes regulated and containing eQTL associated with IMF revealed that those genes were involved in signal transduction, cell cycle, and development pathways, like the AKT signaling (**Figure 5A**). The AKT or PI3K-AKT signaling is an intracellular pathway essential for signal transduction, cell proliferation, apoptosis, and metabolism (Yun et al., 2020). Furthermore, AKT plays a crucial role in adipocyte differentiation. AKT can drive fat production and promote adipogenesis through phosphorylation of substrates, such as Fox family members (Kim et al., 2010; Wang et al., 2020; Yun et al., 2020). Corroborating these findings, in the current study, the *FOXO3* TF, negatively regulated by the 21:22425675 eQTL, was enriched for the Signal transduction AKT signaling. Besides that, although not enriched in this pathway, the *IQGAP1* gene, affected by the same eQTL plays a role in AKT signaling cascades. Li et al. (2017), studying transcriptional differences in pigs in high and low BFT groups, found the PI3K-AKT signaling pathway enriched for differentially expressed liver miRNAs in these animals. Liang et al. (2017), also found the PI3K-AKT pathway related to lipid metabolism and milk fat formation in Holstein cows. These findings indicate that the genes being regulated/containing eQTL associated with IMF participate in relevant lipid-metabolism pathways.

Backfat Thickness

Regarding the eQTL harbored on relevant QTL associated with BFT, most of them were cis-eQTL. Among the affected genes, *EBF4* and *ZSCAN2* are part of the list of bovine curated TFs (de Souza et al., 2018). The first one, *EBF4*, is positively regulated by a single cis-eQTL (rs378953520) harbored on the *PCED1A* gene (**Figure 4B**), and located on the QTL region of BTA13 at 52 Mb (0.22% Vg). This gene is a helix-loop-helix TF, member of the

early B cell factor (Ebf) gene family, that in vertebrates is composed of four members, *EBF1* to *4*. The *EBF1* and *2* participate in the adipogenesis process by playing critical roles during the transcriptional adipogenic cascade (Akerblad et al., 2002; Jimenez et al., 2007). Recently, Cao et al. (2021), in a genome-wide DNA methylation study associated with body fat traits in healthy adult humans, identified a differential methylated position associated with body mass index in the 3'UTR region of the *EBF4* gene, suggesting that this gene could be a target for future obesity risk research. About the gene harboring this mutation, the *PCED1A* is an esterase part of the GDSL/SGNH superfamily, and is expressed in multiple tissues (Maynard et al., 2018). Despite limited studies about the specific functions of this gene, Maynard et al. (2018) indicated a potential structural function at the cell membrane and/or the ECM. As already mentioned, structural genes may exert a function in the cellular rearrangement during adipocyte expansion.

Concerning *ZSCAN2,* this TF is negatively regulated by a cis-eQTL (rs476932264) in the *FSD2* gene (**Figure 4A**), located on BTA21 at 23 Mb, a region that explained 0.21% of Vg in GWAS analysis. Interestingly, *ZSCAN2* expression levels were also negatively regulated by a missense cis-eQTL associated with IMF (**Figure 3A**). This gene belongs to the zinc-finger family of TFs, exerting a role during adipocytes differentiation and determination, and may promote and inhibit PPARγ and C/EBPs expression (Wei et al., 2013; Cassandri et al., 2017). Regarding *FSD2*, this gene not only harbors the rs476932264 cis-eQTL, but also the rs109974605 and rs379905460 (**Figure 4A**). The rs476932264 and rs109974605 are part of the five cis-eQTL affecting their own gene expression levels. While, rs379905460 is a trans-eQTL affecting the expression of both *LOC522763,* a cattle novel gene, and *HAPLN3,* a hyaluronan and proteoglycan-binding link protein gene involved in integrity maintenance and binding functions of the ECM (Spicer et al., 2003). In previous work from our group, a SNP in the *FSD2* gene was already associated with meat color in this Nellore cattle population (Tizioto et al., 2013). Moreover, Lim et al. (2016), indicated it as a potential determinant of overall meat quality in pigs. The authors tested the association of haplotypes produced by *FSD2* SNP and meat quality traits in Berkshire pigs, showing significant associations of the haplotypes with moisture, crude protein levels, color, and IMF content.

We also highlight here the genes *ENSBTAG00000025383* and *NDUFC1*. These two genes were affected by 8 and 9 eQTL associated with BFT, respectively, being the most regulated ones. Although there is a scarcity of information about the novel gene *ENSBTAG00000025383*, it is part of the *NDUFC1* Gene Tree, according to the Ensembl database. The NADH-ubiquinone oxidoreductase (NDUF) enzymes are components of the Complex I oxidative phosphorylation system in mitochondria. In mammals, almost all the ATP molecules required by the cells are generated by oxidative phosphorylation in the mitochondrial respiratory chain (Papa et al., 2002). Kim et al. (2009) and Karisa et al. (2013) found NDUF genes related to beef cattle growth and fat deposition traits (ribeye area and marbling). Corroborating these findings, Jeong et al. (2013) studied transcriptome alterations on the skeletal muscle of castrated Korean cattle that drives IMF deposition and found upregulated NDUF genes enriched for the oxidative phosphorylation process. Cesar et al. (2016), working with this Nellore population, identified NDUF genes differentially expressed in the skeletal muscle associated with fatty acid content. Even though none of the cited studies found these genes associated with the BFT, we have enough evidence to support them as candidate genes.

Lastly, the enrichment analysis of the genes regulated/containing eQTL encompassing relevant QTL associated with the backfat thickness revealed cell cycle, development, and transport pathways. We highlighted the WNT signaling, an important regulator of adipogenesis (Bowers and Lane, 2008). This pathway regulates mesenchymal stem cells, promotes osteogenesis and myogenesis, and inhibits adipogenesis through deacetylation of PPARγ and C/EBPα promoters, and also, by blocking their expression (Bowers and Lane, 2008). In a previous work (Silva-Vignato et al., 2017), studying the skeletal muscle transcriptome of a subset of this Nellore population with extreme values for BFT (high and low groups), we found the WNT signaling enriched for an upregulated gene in the low BFT group. Similarly, Li et al. (2017), working with a pig population with divergent BFT phenotypes (high and low groups), identified the WNT signaling enriched for the differentially expressed miRNAs.

In conclusion, combining RNA-Seq information (expression and SNP) with a high-density genotyping panel, allowed us to identify relevant genomic regions and regulatory polymorphisms associated with intramuscular fat and backfat thickness of Nellore cattle. Within the genes regulated by eQTL associated with the interest traits, we highlight that the transcription factors *FOXP4, FOXO3, ZSCAN2,* and *EBF4* are involved in lipid metabolism-related pathways and may regulate major adipogenesis genes, such as the PPARγ and C/EBPα. We also reported for the first time, a missense cis-eQTL in the *WDR73* gene associated with the intramuscular fat content. These findings help us to improve our knowledge about the genetic architecture that controls economically important carcass and meat quality fat traits in bovine.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee Guidelines from EMBRAPA (CEUA 01/2013).

AUTHOR CONTRIBUTIONS

BS-V, AC, MP, GM, and LLC participated in the experimental design. BS-V, AC, JA, and JP performed data analysis. BS-V

drafted the manuscript. BS-V, AC, JA, MP, GM, JP, IG, LC, GBM, LR, and LLC collaborated with the interpretation and discussion of the results. LLC, GBM, and LR provided the experimental environment and coordination. All authors have read and approved the final manuscript.

FUNDING

The funding was provided by FAPESP (São Paulo Research Foundation, process number 2019/04089-2). LLC, GBM, and LR were granted CNPq (Brazilian National Council for Scientific and Technological Development) fellowships.

ACKNOWLEDGMENTS

The authors thank EMBRAPA and the University of São Paulo for their collaborative efforts. They also thank FAPESP for the scholarship to BS-V (FAPESP process number 2019/18385-2) and for providing financial support (FAPESP process number 2019/04089-2) and CNPq for LLC, GBM, and LR fellowships.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.935238/full#supplementary-material

REFERENCES

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* 31 (2), 166–169. doi:10.1093/bioinformatics/btu638

AOCS (2004). "Rapid Determination of Oil/fat Utilizing High Temperature Solvent Extraction,". AOCS Official Procedure Am 5-04 in *Official Methods and Recommended Practices of the AOCS* (Champaign, IL: American Oil Chemists Society, AOCS).

Akerblad, P., Lind, U., Liberg, D., Bamberg, K., and Sigvardsson, M. (2002). Early B-Cell Factor (O/E-1) Is a Promoter of Adipogenesis and Involved in Control of Genes Important for Terminal Adipocyte Differentiation. *Mol. Cell. Biol.* 22 (22), 8015–8025. doi:10.1128/mcb.22.22.8015-8025.2002

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 Methodol.* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Borges, B. O., Curi, R. A., Baldi, F., Feitosa, F. L. B., Andrade, W. B. F. d., Albuquerque, L. G. d., et al. (2014). Polymorphisms in Candidate Genes and Their Association with Carcass Traits and Meat Quality in Nellore Cattle. *Pesq. Agropec. Bras.* 49 (5), 364–371. doi:10.1590/S0100-204X2014000500006

Bowers, R. R., and Lane, M. D. (2008). Wnt Signaling and Adipocyte Lineage Commitment. *Cell. Cycle* 7 (9), 1191–1196. doi:10.4161/cc.7.9.5815

Brouard, J.-S., Schenkel, F., Marete, A., and Bissonnette, N. (2019). The GATK Joint Genotyping Workflow Is Appropriate for Calling Variants in RNA-Seq Experiments. *J. Anim. Sci. Biotechnol.* 10. doi:10.1186/s40104-019-0359-0

Burrow, H. M., Moore, S. S., Johnston, D. J., Barendse, W., and Bindon, B. M. (2001). Quantitative and Molecular Genetic Influences on Properties of Beef: a Review. *Aust. J. Exp. Agric.* 41 (7), 893–919. doi:10.1071/EA00015

Cai, Z., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2019). Weighting Sequence Variants Based on Their Annotation Increases the Power of Genome-wide Association Studies in Dairy Cattle. *Genet. Sel. Evol.* 51 (1). doi:10.1186/s12711-019-0463-9

Calkhoven, C. F., and Ab, G. (1996). Multiple Steps in the Regulation of Transcription-Factor Level and Activity. *Biochem. J.* 317 (Pt 2), 329–342. doi:10.1042/bj3170329

Cao, V. T., Lea, R. A., Sutherland, H. G., Benton, M. C., Pishva, R. S., Haupt, L. M., et al. (2021). A Genome-wide Methylation Study of Body Fat Traits in the Norfolk Island Isolate. *Nutr. Metabolism Cardiovasc. Dis.* 31 (5), 1556–1563. doi:10.1016/j.numecd.2021.01.027

Cao, Y., Jin, H.-G., Ma, H.-H., and Zhao, Z.-H. (2017). Comparative Analysis on Genome-wide DNA Methylation in Longissimus Dorsi Muscle between Small Tailed Han and Dorper×Small Tailed Han Crossbred Sheep. *Asian-Australas J. Anim. Sci.* 30 (11), 1529–1539. doi:10.5713/ajas.17.0154

Casper, J., Zweig, A. S., Villarreal, C., Tyner, C., Speir, M. L., Rosenbloom, K. R., et al. (2018). The UCSC Genome Browser Database: 2018 Update. *Nucleic Acids Res.* 46 (Database issue), D762–D769. doi:10.1093/nar/gkx1020

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., et al. (2017). Zinc-finger Proteins in Health and Disease. *Cell. Death Discov.* 3, 17071. doi:10.1038/cddiscovery.2017.71

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., Mourão, G. B., Tullio, R. R., Lanna, D. P., 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

de Souza, M. M., Zerlotini, A., Geistlinger, L., Tizioto, P. C., Taylor, J. F., Rocha, M. I. P., et al. (2018). A Comprehensive Manually-Curated Compendium of Bovine Transcription Factors. *Sci. Rep.* 8 (1), 13747. doi:10.1038/s41598-018-32146-2

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-Seq Reads with STAR. *Curr. Protoc. Bioinforma.* 51, 11.19.1–11.14.19. doi:10.1002/0471250953.bi1114s51

Erickson, H. L., and Anakk, S. (2018). Identification of IQ Motif-Containing GTPase-Activating Protein 1 as a Regulator of Long-Term Ketosis. *JCI Insight* 3 (21), e99866. doi:10.1172/jci.insight.99866

Garrick, D. J., and Fernando, R. L. (2013). Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology. *Methods Mol. Biol.* 1019, 275–298. doi:10.1007/978-1-62703-447-0_11

Golson, M. L., and Kaestner, K. H. (2016). Fox Transcription Factors: From Development to Disease. *Development* 143 (24), 4558–4570. doi:10.1242/dev.112672

Gomez-Santos, C., Hernandez-Morante, J. J., Margareto, J., Larrarte, E., Formiguera, X., Martínez, C. M., et al. (2011). Profile of Adipose Tissue Gene Expression in Premenopausal and Postmenopausal Women. *Menopause* 18 (6), 675–684. doi:10.1097/gme.0b013e31820641da

Gross, D. N., van den Heuvel, A. P. J., and Birnbaum, M. J. (2008). The Role of FoxO in the Regulation of Metabolism. *Oncogene* 27, 2320–2336. doi:10.1038/onc.2008.25

Guermah, M., Ge, K., Chiang, C.-M., and Roeder, R. G. (2003). The TBN Protein, Which Is Essential for Early Embryonic Mouse Development, Is an Inducible TAFII Implicated in Adipogenesis. *Mol. Cell.* 12 (4), 991–1001. doi:10.1016/s1097-2765(03)00396-4

Gui, L., and Jia, J. (2018). Effect of Single Nucleotide Polymorphisms in the UCP3 and FOXO1 Genes on Carcass Quality Traits in Qinchuan Cattle. *J. Anim. Feed Sci.* 27 (4), 301–306. doi:10.22358/jafs/97366/2018

Hedman, A. C., Li, Z., Gorisse, L., Parvathaneni, S., Morgan, C. J., and Sacks, D. B. (2021). IQGAP1 Binds AMPK and is Required for Maximum AMPK Activation. *J. Biol. Chem.* 296, 100075. doi:10.1074/jbc.RA120.016193

Hu, Z.-L., Park, C. A., and Reecy, J. M. (2019). Building a Livestock Genetic and Genomic Information Knowledgebase Through Integrative Developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47 (Database issue), D701–D710. doi:10.1093/nar/gky1084

Ishihara, Y., Tsuji, M., and Vogel, C. F. A. (2018). Suppressive Effects of Aryl-Hydrocarbon Receptor Repressor on Adipocyte Differentiation in 3T3-L1 Cells. *Archives Biochem. biophysics* 642, 75–80. doi:10.1016/j.abb.2018.01.018

Jeong, J., Bong, J., Kim, G. D., Joo, S. T., Lee, H.-J., and Baik, M. (2013). Transcriptome Changes Favoring Intramuscular Fat Deposition in the Longissimus Muscle Following Castration of Bulls1. *J. Anim. Sci.* 91 (10), 4692–4704. doi:10.2527/jas.2012-6089

Jiang, C., Gai, N., Zou, Y., Zheng, Y., Ma, R., Wei, X., et al. (2017). WDR73 Missense Mutation Causes Infantile Onset Intellectual Disability and Cerebellar Hypoplasia in a Consanguineous Family. *Clin. Chim. Acta* 464, 24–29. doi:10.1016/j.cca.2016.10.029

Jimenez, M. A., A˚kerblad, P., Sigvardsson, M., and Rosen, E. D. (2007). Critical Role for Ebf1 and Ebf2 in the Adipogenic Transcriptional Cascade. *Mol. Cell. Biol.* 27 (2), 743–757. doi:10.1128/MCB.01557-06

Fernandes Júnior, G. A., Costa, R. B., de Camargo, G. M. F., Carvalheiro, R., Rosa, G. J. M., Baldi, F., et al. (2016). Genome Scan for Postmortem Carcass Traits in Nellore Cattle1. *J. Anim. Sci.* 94 (10), 4087–4095. doi:10.2527/jas.2016-0632

Karisa, B. K., Thomson, J., Wang, Z., Bruce, H. L., Plastow, G. S., and Moore, S. S. (2013). Candidate Genes and Biological Pathways Associated with Carcass Quality Traits in Beef Cattle. *Can. J. Anim. Sci.* 93, 295–306. doi:10.4141/cjas2012-136

Kauffman, R. G., and Berg, E. P. (2011). "Body Composition: Linear Dimensions," in *Encyclopedia of Animal Science*. Editors D. E. Ullrey, C. K. Baer, and W. G. Pond (Boca Raton, FL, USA: Taylor & Francis Group). doi:10.1081/e-eas2-120045651

Kelwick, R., Desanlis, I., Wheeler, G. N., and Edwards, D. R. (2015). The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin Motifs) Family. *Genome Biol.* 16 (1), 113. doi:10.1186/s13059-015-0676-3

Kim, N.-K., Cho, Y.-M., Jung, Y.-S., Kim, G.-S., Heo, K.-N., Lee, S.-H., et al. (2009). Gene Expression Profiling of Metabolism-Related Genes between Top Round and Loin Muscle of Korean Cattle (Hanwoo). *J. Agric. Food Chem.* 57 (22), 10898–10903. doi:10.1021/jf902276r

Kim, S.-p., Ha, J. M., Yun, S. J., Kim, E. K., Chung, S. W., Hong, K. W., et al. (2010). Transcriptional Activation of Peroxisome Proliferator-Activated Receptor-γ Requires Activation of Both Protein Kinase A and Akt during Adipocyte Differentiation. *Biochem. Biophysical Res. Commun.* 399 (1), 55–59. doi:10.1016/j.bbrc.2010.07.038

Leal-Gutiérrez, J. D., Elzo, M. A., Carr, C., and Mateescu, R. G. (2020). RNA-Seq Analysis Identifies Cytoskeletal Structural Genes and Pathways for Meat Quality in Beef. *PLoS ONE* 15, e0240895. doi:10.1371/journal.pone.0240895

Lee, C. (2018). Genome-Wide Expression Quantitative Trait Loci Analysis Using Mixed Models. *Front. Genet.* 9. doi:10.3389/fgene.2018.00341

Lee, S.-H., Gondro, C., van der Werf, J., Kim, N.-K., Lim, D.-j., Park, E.-W., et al. (2010). Use of a Bovine Genome Array to Identify New Biological Pathways for Beef Marbling in Hanwoo (Korean Cattle). *BMC Genomics* 11, 623. doi:10.1186/1471-2164-11-623

Lefterova, M. I., Haakonsson, A. K., Lazar, M. A., and Mandrup, S. (2014). PPARγ and the Global Map of Adipogenesis and beyond. *Trends Endocrinol. Metabolism* 25 (6), 293–302. doi:10.1016/j.tem.2014.04.001

Li, H. (2011). A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data. *Bioinformatics* 27 (21), 2987–2993. doi:10.1093/bioinformatics/btr509

Li, W., Yang, Y., Liu, Y., Liu, S., Li, X., Wang, Y., et al. (2017). Integrated Analysis of mRNA and miRNA Expression Profiles in Livers of Yimeng Black Pigs with Extreme Phenotypes for Backfat Thickness. *Oncotarget* 8 (70), 114787–114800. doi:10.18632/oncotarget.21918

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 (1), 6396. doi:10.1038/s41598-017-06634-w

Lim, K.-S., Lee, K.-T., Lee, S.-W., Chai, H.-H., Jang, G., Hong, K.-C., et al. (2016). Genomic Structure, Expression and Association Study of the Porcine FSD2. *Mol. Biol. Rep.* 43 (9), 1011–1018. doi:10.1007/s11033-016-4029-4

Littlejohn, M. D., Tiplady, K., Fink, T. A., Lehnert, K., Lopdell, T., Johnson, T., et al. (2016). Sequence-based Association Analysis Reveals an MGST1 eQTL with Pleiotropic Effects on Bovine Milk Composition. *Sci. Rep.* 6, 25376. doi:10.1038/srep25376

Liu, J., Nong, Q., Wang, J., Chen, W., Xu, Z., You, W., et al. (2020). Breed Difference and Regulatory Role of CRTC3 in Porcine Intramuscular Adipocyte. *Anim. Genet.* 51 (4), 521–530. doi:10.1111/age.12945

Liu, J., Wang, L., Chen, W., Li, J., and Shan, T. (2021). CRTC3 Regulates the Lipid Metabolism and Adipogenic Differentiation of Porcine Intramuscular and Subcutaneous Adipocytes by Activating the Calcium Pathway. *J. Agric. Food Chem.* 69 (25), 7243–7255. doi:10.1021/acs.jafc.1c02021

Lopes, L. S., Ladeira, M. M., Machado Neto, O. R., Paulino, P. V. R., Chizzotti, M. L., Ramos, E. M., et al. (2012). Características de carcaça e cortes comerciais de tourinhos Red Norte e Nelore terminados em confinamento. *R. Bras. Zootec.* 41 (4), 970–977. doi:10.1590/S1516-35982012000400020

Lu, Y.-H., Dallner, O. S., Birsoy, K., Fayzikhodjaeva, G., and Friedman, J. M. (2015). Nuclear Factor-Y Is an Adipogenic Factor that Regulates Leptin Gene Expression. *Mol. Metab.* 4 (5), 392–405. doi:10.1016/j.molmet.2015.02.002

Martins, R., Brito, L. F., Machado, P. C., Pinto, L. F. B., Silva, M. R., Schenkel, F. S., et al. (2021). Genome-wide Association Study and Pathway Analysis for Carcass Fatness in Nellore Cattle Measured by Ultrasound. *Anim. Genet.* 52, 730–733. doi:10.1111/age.13129

Maynard, R. D., Godfrey, D. A., Medina-Gomez, C., and Ackert-Bicknell, C. L. (2018). Characterization of Expression and Alternative Splicing of the Gene Cadherin-like and PC Esterase Domain Containing 1 (Cped1). *Gene* 674, 127–133. doi:10.1016/j.gene.2018.06.060

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17 (1), 122. doi:10.1186/s13059-016-0974-4

Michaelson, J. J., Loguercio, S., and Beyer, A. (2009). Detection and Interpretation of Expression Quantitative Trait Loci (eQTL). *Methods* 48 (3), 265–276. doi:10.1016/j.ymeth.2009.03.004

Moreira, G. C. M., Boschiero, C., Cesar, A. S. M., Reecy, J. M., Godoy, T. F., Trevisoli, P. A., et al. (2018). A Genome-wide Association Study Reveals Novel Genomic Regions and Positional Candidate Genes for Fat Deposition in Broiler Chickens. *BMC genomics* 19 (1), 374. doi:10.1186/s12864-018-4779-6

Mudadu, M. A., Porto-Neto, L. R., Mokry, F. B., Tizioto, P. C., Oliveira, P. S. N., Tullio, R. R., et al. (2016). Genomic Structure and Marker-Derived Gene Networks for Growth and Meat Quality Traits of Brazilian Nelore Beef Cattle. *BMC Genomics* 17, 235. doi:10.1186/s12864-016-2535-3

Mueller, J. C. (2004). Linkage Disequilibrium for Different Scales and Applications. *Briefings Bioinforma.* 5 (4), 355–364. doi:10.1093/bib/5.4.355

Onteru, S. K., Gorbach, D. M., Young, J. M., Garrick, D. J., Dekkers, J. C. M., and Rothschild, M. F. (2013). Whole Genome Association Studies of Residual Feed Intake and Related Traits in the Pig. *PLoS ONE* 8 (6), e61756. doi:10.1371/journal.pone.0061756

Papa, S., Sardanelli, A. M., Scacco, S., Petruzzella, V., Technikova-Dobrova, Z., Vergari, R., et al. (2002). Minireview: The NADH: Ubiquinone Oxidoreductase (Complex I) of the Mammalian Respiratory Chain and the cAMP Cascade. *J. Bioenergetics Biomembr.* 34, 1–10. doi:10.1023/a:1013863018115

Park, S. J., Beak, S.-H., Jung, D. J. S., Kim, S. Y., Jeong, I. H., Piao, M. Y., et al. (2018). Genetic, Management, and Nutritional Factors Affecting Intramuscular Fat Deposition in Beef Cattle - A Review. *Asian-Australas J. Anim. Sci.* 31 (7), 1043–1061. doi:10.5713/ajas.18.0310

Plekhanova, E., Nuzhdin, S. V., Utkin, L. V., and Samsonova, M. G. (2018). Prediction of Deleterious Mutations in Coding Regions of Mammals with Transfer Learning. *Evol. Appl.* 12 (1), 18–28. doi:10.1111/eva.12607

Poleti, M. D., Regitano, L. C. A., Souza, G. H. M. F., Cesar, A. S. M., Simas, R. C., Silva-Vignato, B., et al. (2018). Longissimus Dorsi Muscle Label-free Quantitative Proteomic Reveals Biological Mechanisms Associated with Intramuscular Fat Deposition. *J. Proteomics* 179, 30–41. doi:10.1016/j.jprot.2018.02.028

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033

Raza, S. H. A., Khan, S., Amjadi, M., Abdelnour, S. A., Ohran, H., Alanazi, K. M., et al. (2020). Genome-wide Association Studies Reveal Novel Loci Associated with Carcass and Body Measures in Beef Cattle. *Archives Biochem. Biophysics* 694, 108543. doi:10.1016/j.abb.2020.108543

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

Shabalin, A. A. (2012). Matrix eQTL: Ultra Fast eQTL Analysis via Large Matrix Operations. *Bioinformatics* 28 (10), 1353–1358. doi:10.1093/bioinformatics/bts163

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 (11), 2498–2504. doi:10.1101/gr.1239303

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 (1), 506. doi:10.1186/s12864-017-3897-x

Silva-Vignato, B., Coutinho, L. L., Poleti, M. D., Cesar, A. S. M., Moncau, C. T., Regitano, L. C. A., et al. (2019). Gene Co-expression Networks Associated with Carcass Traits Reveal New Pathways for Muscle and Fat Deposition in Nelore Cattle. *BMC Genomics* 20 (1), 32. doi:10.1186/s12864-018-5345-y

Spicer, A. P., Joo, A., and Bowling, R. A., Jr. (2003). A Hyaluronan Binding Link Protein Gene Family Whose Members Are Physically Linked Adjacent to Chrondroitin Sulfate Proteoglycan Core Protein Genes. *J. Biol. Chem.* 278 (23), 21083–21091. doi:10.1074/jbc.M213100200

Spielman, R. S., Bastone, L. A., Burdick, J. T., Morley, M., Ewens, W. J., and Cheung, V. G. (2007). Common Genetic Variants Account for Differences in Gene Expression Among Ethnic Groups. *Nat. Genet.* 39 (2), 226–231. doi:10.1038/ng1955

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. Bioinforma.* 54, 1.33.1–1.130.33. doi:10.1002/cpbi.5

Suárez-Vega, A., Gutiérrez-Gil, B., Benavides, J., Perez, V., Tosser-Klopp, G., Klopp, C., et al. (2015). Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PloS one* 10 (5), e0126416. doi:10.1371/journal.pone.0126416

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and Limitations of Genome-wide Association Studies. *Nat. Rev. Genet.* 20 (8), 467–484. doi:10.1038/s41576-019-0127-1

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., 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 (21), gkv711. 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 (21), 1012–1020. doi:10.1152/physiolgenomics.00066.2013

Troy, D. J., Tiwari, B. K., and Joo, S.-T. (2016). Health Implications of Beef Intramuscular Fat Consumption. *Korean J. Food Sci. Animal Resour.* 36 (5), 577–582. doi:10.5851/kosfa.2016.36.5.577

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 Strien, J. (2018). *Analysis of eQTL Data Using Machine Learning to Investigate the Potential of Coding Variants to Affect Gene Expression* (Nijmegen, Holland: external research internship Radboud University). PhD Msc Thesis.

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 (4), 252–263. doi:10.1038/nrg2538

Wang, H., Zheng, Y., Wang, G., and Li, H. (2013). Identification of microRNA and Bioinformatics Target Gene Analysis in Beef Cattle Intramuscular Fat and Subcutaneous Fat. *Mol. Biosyst.* 9 (8), 2154–2162. doi:10.1039/c3mb70084d

Wang, L., Zhang, S., Cheng, G., Mei, C., Li, S., Zhang, W., et al. (2020). MiR-145 Reduces the Activity of PI3K/Akt and MAPK Signaling Pathways and Inhibits Adipogenesis in Bovine Preadipocytes. *Genomics* 112, 2688–2694. doi:10.1016/j.ygeno.2020.02.020

Wei, S., Zhang, L., Zhou, X., Du, M., Jiang, Z., Hausman, G. J., et al. (2013). Emerging Roles of Zinc Finger Proteins in Regulating Adipogenesis. *Cell. Mol. Life Sci.* 70 (23), 4569–4584. doi:10.1007/s00018-013-1395-0

Westra, H.-J., and Franke, L. (2014). From Genome to Function by Studying eQTLs. *Biochimica Biophysica Acta (BBA) - Mol. Basis Dis.* 1842 (10), 1896–1902. doi:10.1016/j.bbadis.2014.04.024

Yokoo, M. J., Albuquerque, L. G., Lôbo, R. B., Bezerra, L. A. F., Araujo, F. R. C., Silva, J. A. V., et al. (2008). Genetic and Environmental Factors Affecting Ultrasound Measures of Longissimus Muscle Area and Backfat Thickness in Nelore Cattle. *Livest. Sci.* 117 (Issues 2–3), 147–154. doi:10.1016/j.livsci.2007.12.006

Yokoo, M. J., Lobo, R. B., Araujo, F. R. C., Bezerra, L. A. F., Sainz, R. D., and Albuquerque, L. G. (2010). Genetic Associations between Carcass Traits Measured by Real-Time Ultrasound and Scrotal Circumference and Growth Traits in Nelore Cattle1. *J. Anim. Sci.* 88 (1), 52–58. doi:10.2527/jas.2008-1028

Yun, J., Yu, Y., Zhou, G., Luo, X., Jin, H., Zhao, Y., et al. (2020). Effects of Puerarin on the Akt Signaling Pathway in Bovine Preadipocyte Differentiation. *Asian-Australas J. Anim. Sci.* 33 (1), 4–11. doi:10.5713/ajas.19.0004

Zhbannikov, I. Y., Hunter, S. S., Foster, J. A., and Settles, M. L. (2017). "SeqyClean: A Pipeline for High-Throughput Sequence Data Preprocessing," in Unpublished paper presented at the *8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (Boston, MA: Association for Computing Machinery Digital Library).

Zhu, H. (2016). Forkhead Box Transcription Factors in Embryonic Heart Development and Congenital Heart Disease. *Life Sci.* 144, 194–201. doi:10.1016/j.lfs.2015.12.001

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 19 October 2022
DOI 10.3389/fvets.2022.1031786

OPEN ACCESS

EDITED BY
Ana Fabrícia Braga Magalhães, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

REVIEWED BY
Marta Serna García, Universidad Europea de Valencia, Spain
Sayed Haidar Abbas Raza, Northwest A&F University, China

*CORRESPONDENCE
Shenglong Wu
pigbreeding@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

RECEIVED 30 August 2022
ACCEPTED 05 October 2022
PUBLISHED 19 October 2022

CITATION
Shi J, Xu C, Wu Z, Bao W and Wu S (2022) Integrated analysis of lncRNA-mediated ceRNA network involved in immune regulation in the spleen of Meishan piglets. *Front. Vet. Sci.* 9:1031786. doi: 10.3389/fvets.2022.1031786

Integrated analysis of lncRNA-mediated ceRNA network involved in immune regulation in the spleen of Meishan piglets

Jing Shi[1†], Chao Xu[1†], Zhengchang Wu[1], Wenbin Bao[1,2] and Shenglong Wu[1,2*]

[1]Key Laboratory for Animal Genetics, Breeding, Reproduction and Molecular Design of Jiangsu Province, College of Animal Science and Technology, Yangzhou University, Yangzhou, China, [2]Joint International Research Laboratory of Agriculture and Agri-Product Safety, Yangzhou University, Yangzhou, China

Meishan pigs are a famous local pig breed in China, with high fertility and early sexual maturity, and stronger immunity compared to other breeds. The spleen is the largest lymphoid organ in pigs and performs essential functions, such as those relating to immunity and haematopoiesis. The invasion of many pathogenic microorganisms in pigs is associated with spleen damage. Long non-coding RNAs participate in a broad range of biological processes and have been demonstrated to be associated with splenic immune regulation. However, the expression network of mRNAs and lncRNAs in the spleen of Meishan pigs remains unclear. This study collected spleen tissues from Meishan piglets at three different ages as a model, and mRNA and lncRNA transcripts were profiled for each sample. Additionally, 1,806 differential mRNAs and 319 differential lncRNAs were identified. A complicated interaction between mRNAs and lncRNAs was identified *via* WGCNA, demonstrating that lncRNAs are a crucial regulatory component in mRNA. The results show that the modules black and red have similar mRNA and lncRNA transcription patterns and are mainly involved in the process of the immune defense response. The core genes (DHX58 and IFIT1) and key lncRNAs (TCONS-00002102 and TCONS-00012474) of piglet spleen tissue were screened using the ceRNA network. The expression of these genes is related to the immune response of pigs. Our research may contribute to a further understanding of mRNA and lncRNA expression in the spleen of piglets, and provide new ideas to improve the disease resistance of piglets.

KEYWORDS

Meishan piglet, spleen, mRNA, lncRNA, WGCNA, ceRNA

Introduction

With the rapid development of high-throughput, transcriptome-sequencing technologies, our understanding of the mammalian genome is growing rapidly. This has contributed to the improvement of economic traits in mammals, such as beef cattle (1) and Ningxiang pigs (2). Although approximately two-thirds of the mammalian genome is actively transcribed, most does not encode proteins (3). Long non-coding RNA (lncRNA) is a type of non-coding RNA that accounts for more than 80% of non-coding RNAs (4), with fragments larger than 200 bp, mostly without coding ability (very few can encode small peptides), and is directly involved in the regulation of expression in cells (5). A series of recent papers have shown that cis-lncRNA and trans-lncRNA play a role in regulating the activity of genes (6). Moreover, lncRNA can act as a positive or negative signal during transcription, regulate protein activity, and modulate chromatin function (7–9). Numerous studies have also shown that lncRNAs play an essential regulatory role in immune-related processes in pigs, such as physiopathology. For example, Wu et al. identified differentially expressed lncRNAs and mRNAs in porcine alveolar macrophages after infection with PRRSV (Porcine Reproductive and Respiratory Syndrome Virus) and found that co-expressed genes for down-regulated lncRNAs were significantly enriched in NF-κB and Toll-like receptor signaling pathways (10). Chen et al. identified the expression profile of lncRNAs in IPEC-J2 cells during PEDV (Porcine Endemic Diarrhea Virus) infection and further determined the differential expression of immune-related lncRNAs in PEDV-infected IPEC-J2 cells and newborn piglets (11). Fang et al. analyzed 199 differentially expressed lncRNAs in IPEC-J2 cells after PCV2 (Porcine Circovirus Type 2) infection, and their regulatory target genes (SOD2, TNFAIP3, and MG7) were all associated with infectious diseases (12). LncRNAs, especially during transcription, have remained a research hotspot in recent years. Therefore, in the present study, we investigated the mRNA regulation of lncRNAs in Meishan piglets.

China has approximately 100 breeds of domestic pig genetic resources, accounting for over one-third of the global total (13). The Meishan pig is a famous local pig in China, known for its high fecundity and early sexual maturity. Additionally, many domestic and foreign studies have shown that Meishan pigs have shown stronger tolerance and resistance to many diseases. For example, Halbur et al. (14) found that Meishan pigs were less susceptible to Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) than Hampshire and Duroc pigs. Reiner et al. (15) found that Meishan pigs were more resistant to pork sporozoites than Pietrain pigs. In addition, Chen et al. (16) compared the expression of porcine β-defensins in Meishan pigs and crossbred (Duroc × Yorkshire × Landrace) pigs and found that Meishan pigs had a higher expression, which might be the reason for their higher immunity and disease resistance. Dong et al. (17) compared the intestinal barrier function of Meishan piglets and crossbred neonatal piglets and found that Meishan pigs had greater intestinal barrier function. Therefore, it is of great interest to study the immune resistance of Meishan pigs.

Pig farming is a pillar of China's livestock industry, with epidemic problems causing considerable losses to the pig industry every year. In particular, piglets are less resistant to disease than adult pigs, and there are differences between breeds in the resistance of pigs to disease (18). PED (Porcine Epidemic Diarrhea) affects pigs of all ages, but lethality is seen mainly in lactating piglets (19). Newborn piglets infected with PDCoV (Porcine Delta Coronavirus) die from severe diarrhea (20). Infection with TGEV (Transmissible Gastroenteritis Virus) causes 100% mortality in piglets under 14 days of age (21). It is therefore a matter of urgency to pay attention to and improve piglets' immunity and disease resistance. As the most prominent secondary lymphoid organ in pigs, the spleen contains a variety of immunoreactive cells and immune factors. It is an important site of response for both innate and adaptive immunity. In recent years, many studies have been conducted to screen for immune-related genes in pigs by the transcriptome analysis of porcine immune organs in order to improve resistance and resilience to pathogens (22–25).

However, no RNA sequencing study has yet been reported for Meishan pigs. In order to investigate the molecular regulatory mechanisms of immune differences among Meishan piglets, the spleen was used as an immune model. The spleens were collected from Meishan piglets at three ages: 1 day old (without colostrum), 14 days after colostrum feeding, and 28 days after colostrum feeding. Additionally, the mRNA and lncRNA sequencing of spleen tissues was performed by high-throughput sequencing technology. We successfully identified differentially expressed candidate genes, thus providing a reference for studying the function and mechanism of lncRNAs in spleen tissues. In addition, our study will increase our understanding of the transcriptomics associated with spleen tissue and contribute to a better understanding of the immune function of the mammalian spleen.

Materials and methods

Ethics statement

All experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Yangzhou University (Pig: SYXK(Su)2012-0029) and were performed according to the Animal Ethics Procedures and Guidelines of the People's Republic of China. No other specific permissions were required for these experiments.

Experimental animal and sample collection

Two Meishan sows were chosen that were similar in weight, age, and body shape and that had farrowed on the same day. After parturition, the piglets received the same diet and were housed in an environmentally controlled room. We designated the first day of the newborns' existence as day 1. Once the sows farrowed, two piglets from each litter were immediately chosen to be slaughtered. Four piglets of similar weight were sacrificed. Following slaughter, spleen samples were collected simultaneously and snap-frozen in liquid nitrogen (-196°C). The remaining piglets were housed in two pens in an environmentally controlled room and were fed under identical husbandry conditions. All piglets were fed only *via* breastfeeding until weaning (day 35). Similar to the aforementioned sample collection procedure on day 1, two piglets from each litter were chosen to be sacrificed at postnatal days 14 and 28 after farrowing using an intravenous injection of pentobarbital sodium, which minimized animal suffering. Spleen samples were immediately stored in liquid nitrogen for RNA isolation. Samples were kept at ultra-low temperatures to avoid RNA degradation.

RNA extraction

Total RNA from the milled spleen tissues was extracted using Trizol reagent (TaKaRa, Dalian, China), according to the manufacturer's protocol. The extracted total RNA was then treated with RNase-free DNase to remove excess DNA. The quality of the RNA extracted from spleen tissue was assessed by Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Qualified total RNA was stored at −80°C until use. A total of 12 samples were used for RNA extraction.

Library construction and RNA-Seq analysis

Qualified RNA samples of four individuals at the same age were equally pooled together to form three RNA groups: 1 d, 14 d, and 28 d. Approximately 1 μg of total RNA per sample was treated with the Ribo-Zero™ Magnetic Kit (Epicenter) to deplete rRNA. The retrieved RNA was fragmented by adding First Strand Master Mix (Invitrogen). First-strand cDNA was generated using random primer reverse transcription, followed by second-strand cDNA synthesis. The synthesized cDNA was subjected to end-repair and then was 3' adenylated. Adapters were ligated to the ends of these 3' adenylated cDNA fragments. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the cDNA fragments. Then, the PCR products were purified with Ampure XP Beads. The constructed RNA libraries were quality checked with an Agilent 2100 Bioanalyzer and then sequenced using an Illumina sequencer.

Raw data quality assessment

Raw sequencing data (Raw Reads) were first filtered to obtain high-quality clean data to ensure the quality and accuracy of subsequent bioinformatic analysis. The quality control of the raw and trimmed reads was performed using FastQC and MultiQC (26, 27). Trimming of the adapter content and quality trimming was performed using Cutadapt (28). All downstream analyses were based on high-quality clean reads.

Identification and classification of lncRNAs

Clean reads aligned to the reference genome (Sus scrofa 11.1) were stored in a binary bam file. The new transcript was spliced after the readings were collated using Stringtie software (29). Then, by comparing the gene annotation data of the reference sequence generated by Cuffcompare software (30), the potential lncRNA transcripts were chosen. CPC, CNCI, Pfam, and PLEK were used to filter out lncRNAs with coding potential and obtain the predicted lncRNA sequences (31–34).

Functional enrichment and differential expression analysis

Using the R package DESeq2 (35), differential expression gene analysis was conducted between the two groups, and genes with p-adj < 0.05 and $|\log2FoldChange| > 1$ were chosen as differential genes. By performing a hypergeometric distribution test using enrichGO and enrichKEGG in the R package clusterProfiler (36), the functional enrichment analysis of genes based on the GO and KEGG databases was carried out and enriched pathways with $p < 0.05$ were retained.

Co-expression networks (weighted correlation network analysis)

Weighted gene co-expression network analysis (WGCNA) can be used to build gene co-expression modules using gene expression profiles (37, 38). The gene relationship matrix was first derived from the gene expression matrix using the Pearson correlation coefficient. By setting a soft threshold of 9, the gene relationship matrix was transformed into an adjacency matrix.

The network's interconnectivity was then determined using the topological overlap matrix (TOM). In order to categorize genes into various modules, we employed the TOM difference degree as the clustering distance. Using a threshold of 0.25, the dynamic tree approach was also employed to combine related gene modules (37).

Construction of mRNA–lncRNA and mRNA–lncRNA–pathway networks

To explore the association between mRNA and lncRNA using the significant module of the mRNA–lncRNA co-expression network, we constructed mRNA–lncRNA–pathway networks based on mRNA–lncRNA networks and important pathways involved in mRNA regulation. This paper aims to reveal the relevant pathways of lncRNA regulation and thus predict the possible mechanisms of lncRNA in the spleen. The network was constructed using significant correlation pairs based on Pearson correlation coefficients. The differential co-expression network was visualized and analyzed using Cytoscape software (version: 3.9.0) (39).

Establishment of the lncRNA–miRNA–mRNA ceRNA network

In 2011, Leonardo Salmena proposed the competing endogenous RNA (ceRNA) hypothesis, revealing a new mechanism of RNA interactions. This hypothesis suggested that different types of RNA molecules competitively bind to miRNAs, thus reducing the inhibitory effect of miRNAs on their target mRNAs; these competitive endogenous RNAs may include circRNAs, lncRNAs, pseudogenes, and protein-coding mRNAs (40). To better understand the role of lncRNAs in the ceRNA network, all potential co-deregulated competitive triads were established to construct the lncRNA–miRNA–mRNA network. The miRNA sequences were downloaded from the miRBase website. The MiRanda software (41) was used to predict miRNA-targeted lncRNAs and mRNAs. A max score > 190 and max energy < -15 were set to obtain high confidence in the interaction relationships. The ceRNA network was then constructed from co-expressed mRNA–lncRNAs and their co-targeted miRNAs (42).

Real-time PCR quantification

Total RNAs from pig spleen tissues were extracted using Trizol (TaKaRa, Dalian, China). Then, reverse transcription of RNA was conducted using HiScript III RT SuperMix with gDNA wiper (Vazyme, Nanjing, China). The RT-qPCR reaction system contained 5 μl SYBR Green Mixture (Vazyme, Nanjing, China), 1 μl of the cDNA template, 0.2 μl of each primer, and 3.6 μl deionised water. The thermal conditions were as follows: 95°C for 5 min, 40 cycles of 95°C for 10 s, and 60°C for 30 s. The GAPDH genes were set as the internal controls. All the forward and reverse primers for the RT-qPCR assays are listed in Table 1. The expression level of each validated gene for each timepoint was calculated by the $2^{-\Delta\Delta Ct}$ method.

TABLE 1 Forward and reverse primers used for gene quantification by RT-qPCR.

Name	Sequence (5'−3')
IFIT1-F	CTTGGAGGAGATTGAGTT
IFIT1-R	CAGTATGTTCTTGTTGGG
DHX58-F	CTCTGTGCCAACCT
DHX58-R	TCCCGTCTCAACTC
TCONS-00012474-F	GAGCCACAAAGGGAA
TCONS-00012474-R	GCTGAGGTGAGGTAA
TCONS-00002102-F	GCCCTTCTACCCTATCAT
TCONS-00002102-R	ATTTCCTTTCACCGACTC

Results

Identification and classification of lncRNAs in the spleen tissue of Meishan piglets

The data analysis process for this study is shown in Figure 1. We used a combination of the most widely used coding potential analysis methods to screen the candidate lncRNAs, including CPC analysis, CNCI analysis, Pfam protein structural domain analysis, and PLEK analysis. The resulting Venn diagram shows that a total of 2,234 new lncRNA transcripts were detected using the four methods (Figure 2A). The statistical distribution of the GC content of the predicted lncRNA sequences showed that most of the lncRNA GC content was around 30–60% (Figure 2B). The stacked bar chart shows that the highest proportion of newly predicted lncRNAs was genic intronic in both the antisense and sense categories (Figure 2C). Figure 2D shows the length distribution of lncRNAs, with more lncRNAs being between 200 and 300 bp in length (Figure 2D). The results of the exon number distribution map of lncRNAs show that lncRNAs were mainly concentrated in two exons, followed by three exons (Figure 2E).

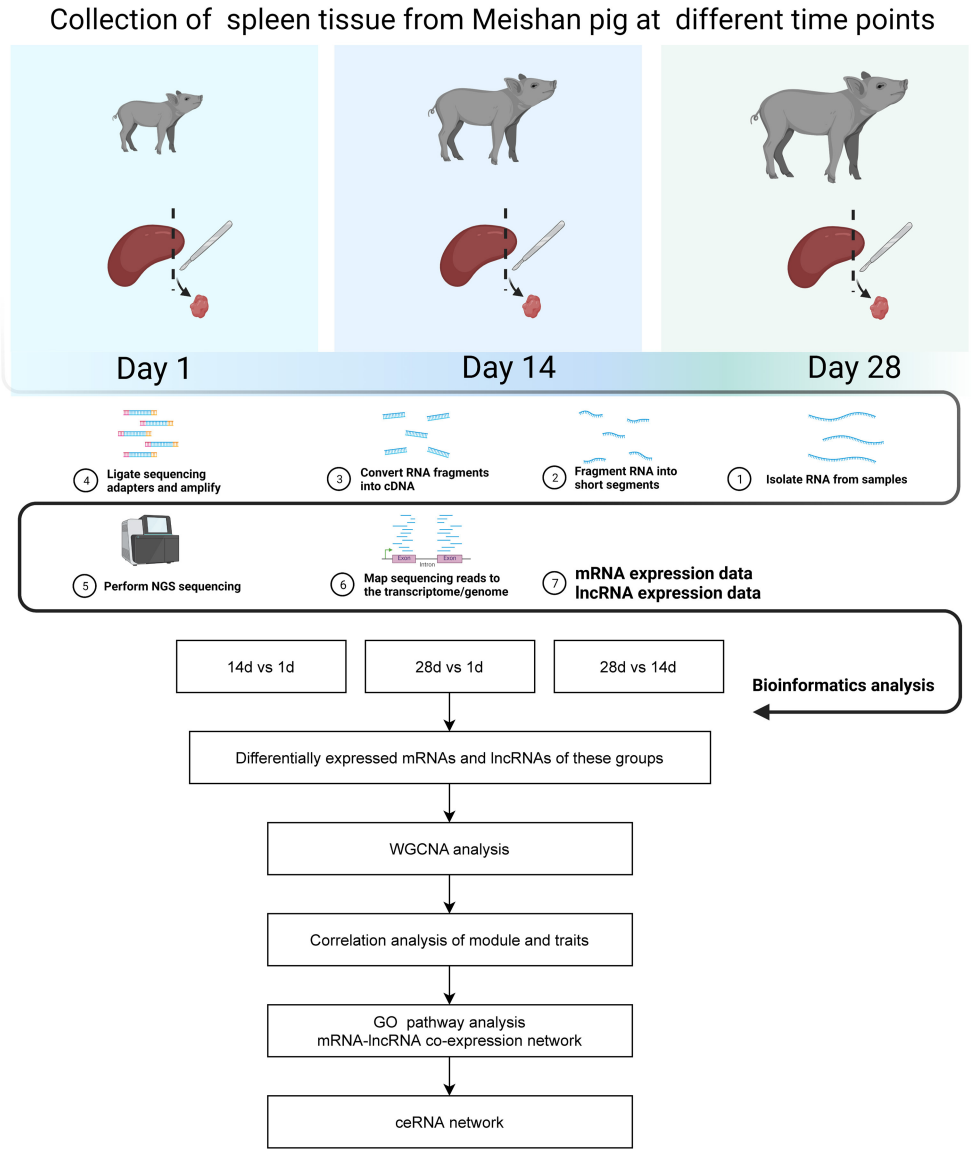


FIGURE 1
Flow chart of data analysis.

Analysis of differentially expressed mRNA and lncRNA

The box plot demonstrates the consistency in the distribution of mRNA and lncRNA in terms of the transcript expression levels across the three groups, indicating no batch effect in the data (Figure 3A). Principal component analysis showed that the three different groups of spleens essentially formed well-defined groups, ranked according to their number of days (Figure 3B). Day 1 and days 14 and 28 clustered more distantly, while days 14 and 28 clustered more closely with each other. Differences in the expression of mRNAs and lncRNAs between the three groups were analyzed by volcano plots, which showed 367 up-regulated genes and 188 down-regulated genes for 14 vs. 1 d, 748 up-regulated genes and 439 down-regulated genes for 28 vs. 1 d, and 51 up-regulated genes and 13 down-regulated genes for 28 vs. 14 d in mRNAs. In terms of lncRNAs, 14 vs. 1 d had 62 up-regulated genes and 21 down-regulated genes, 28 vs. 1 d had 148 up-regulated genes and 73 down-regulated genes, and 28 vs. 14 d had 9 up-regulated genes and 6 down-regulated genes (Figure 3C, Supplementary Tables 1, 2).

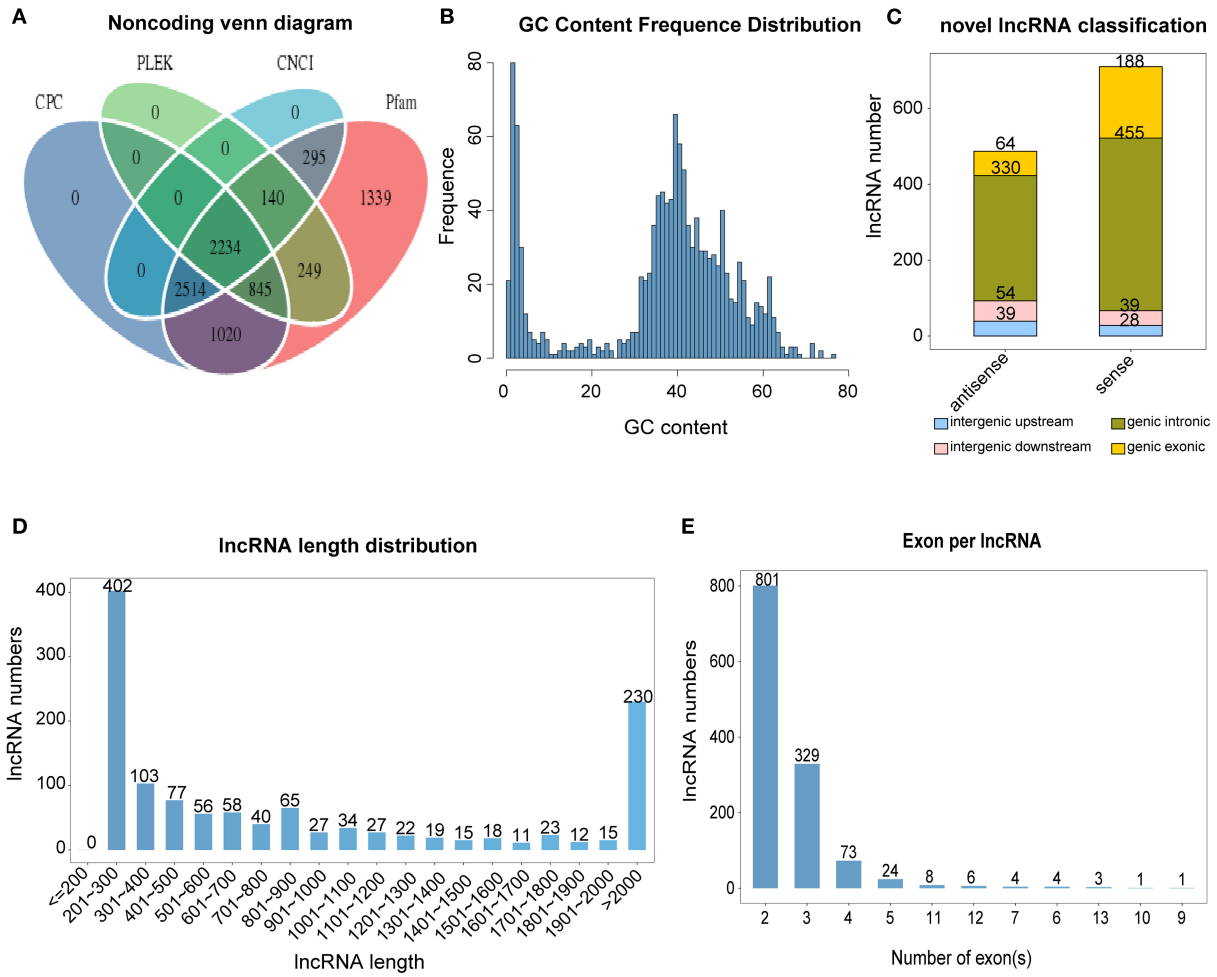


FIGURE 2
Classification and identification of lncRNAs. **(A)** Venn diagram of lncRNA, predicted using CPC (Coding Potential Calculator), CNCI (Coding-Non-Coding Index), CPAT (Coding Potential Assessment Tool), and Pfam (Protein families). **(B)** lncRNA GC content. **(C)** lncRNA classification. **(D)** lncRNA length. **(E)** lncRNA exon number.

GO and KEGG enrichment analysis of differentially expressed mRNAs

GO analysis revealed that the biological processes and pathways enriched by the differentially expressed mRNAs in piglet spleen tissues were mainly related to the immune response, defense response, and defense response to a virus (Figure 4A, Supplementary Tables 3–5). KEGG pathway enrichment analysis showed that cytokine–cytokine receptor interaction, haematopoietic cell lineage, toll-like receptor signaling pathway, and the NF-kappa B signaling pathway were the most important pathways (Figure 4B, Supplementary Tables 6–8).

WGCNA analysis

WGCNA was used to construct a differentially expressed co-expression module comprising 1,806 mRNAs and 319 lncRNAs. When the soft threshold power β was set at 9, the scale-free network matching index exceeded 0.7, which is of greater biological significance (Figure 5A). Therefore, β = 9 was used to generate a hierarchical clustering tree. A co-expression network (combined cut height = 0.25, redundancy = 3) was constructed by WGCNA to discover the potential regulatory functional relationships between lncRNAs and mRNAs in spleen tissues at different ages, as well as their mechanisms. The network was divided into five modules, identified by and displayed

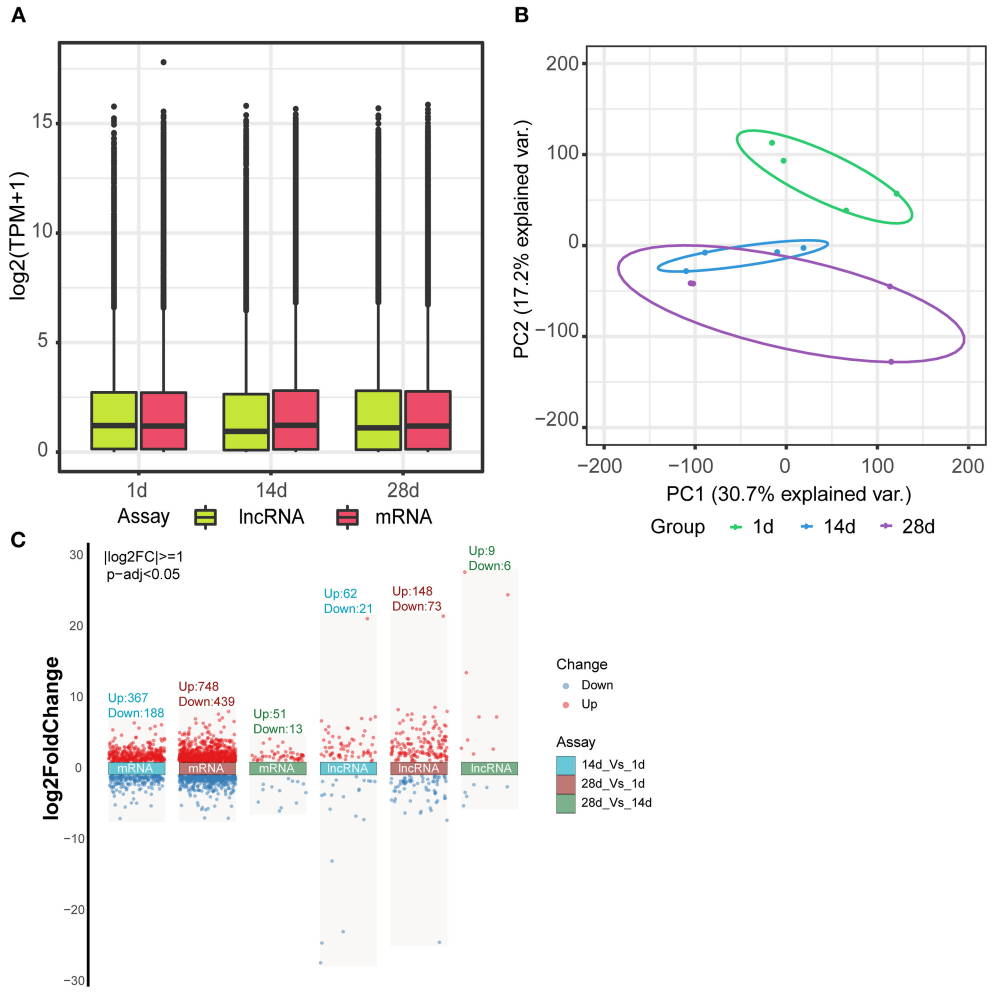


FIGURE 3
Differential expression of mRNAs and lncRNAs. **(A)** Box line plot of mRNA and lncRNA expression in three different groups of spleen tissues. **(B)** Principal component analysis of mRNA and lncRNA in three different groups of spleen tissues. **(C)** Volcano plot analysis of mRNA and lncRNA expression between the three groups.

in different colors (black, blue, yellow, gray, and red), each containing different gene clusters and showing the expression patterns of genes within the different modules in a heat map (Figure 5B, Supplementary Table 9). Red represents positive correlations, while blue represents negative correlations. The relationship between co-expression modules and different ages is shown in Figure 5C. We found a significant positive correlation between the black and red modules and 28 days (correlation coefficient for 28 d = 0.71, $p = 0.01$). Gene significance analysis was performed for each of the 28 d modules (Figure 5D), and the red and black modules were significant > 0.5; thus, the red and black modules for 28 d were selected for the next step of the analysis. Figure 5E shows the significance of these genes in the red and black module for 28 d. Furthermore, genes with gene significance > 0.6 and module membership > 0.8 were used as core genes. There were 69 mRNAs and 24 lncRNAs in the black and red modules (Figure 5E). Performing GO enrichment analysis of mRNAs, we found that the red and black modules are mainly involved in the process of the immune defense response (Figure 5F, Supplementary Table 10).

Construction of mRNA–lncRNA co-expression networks and ceRNA networks

In our study, we identified 14 mRNAs and 2 lncRNAs in the mRNA–lncRNA pathway network, all of which were upregulated at 28 days (Figure 6A). In the mRNA–lncRNA network, *TCONS-00012474* (one lncRNA) was linked to 14 mRNAs (*DLG3, SCG3, CLEC4F, GZMB, IRF9, IFI44, RTP4,*

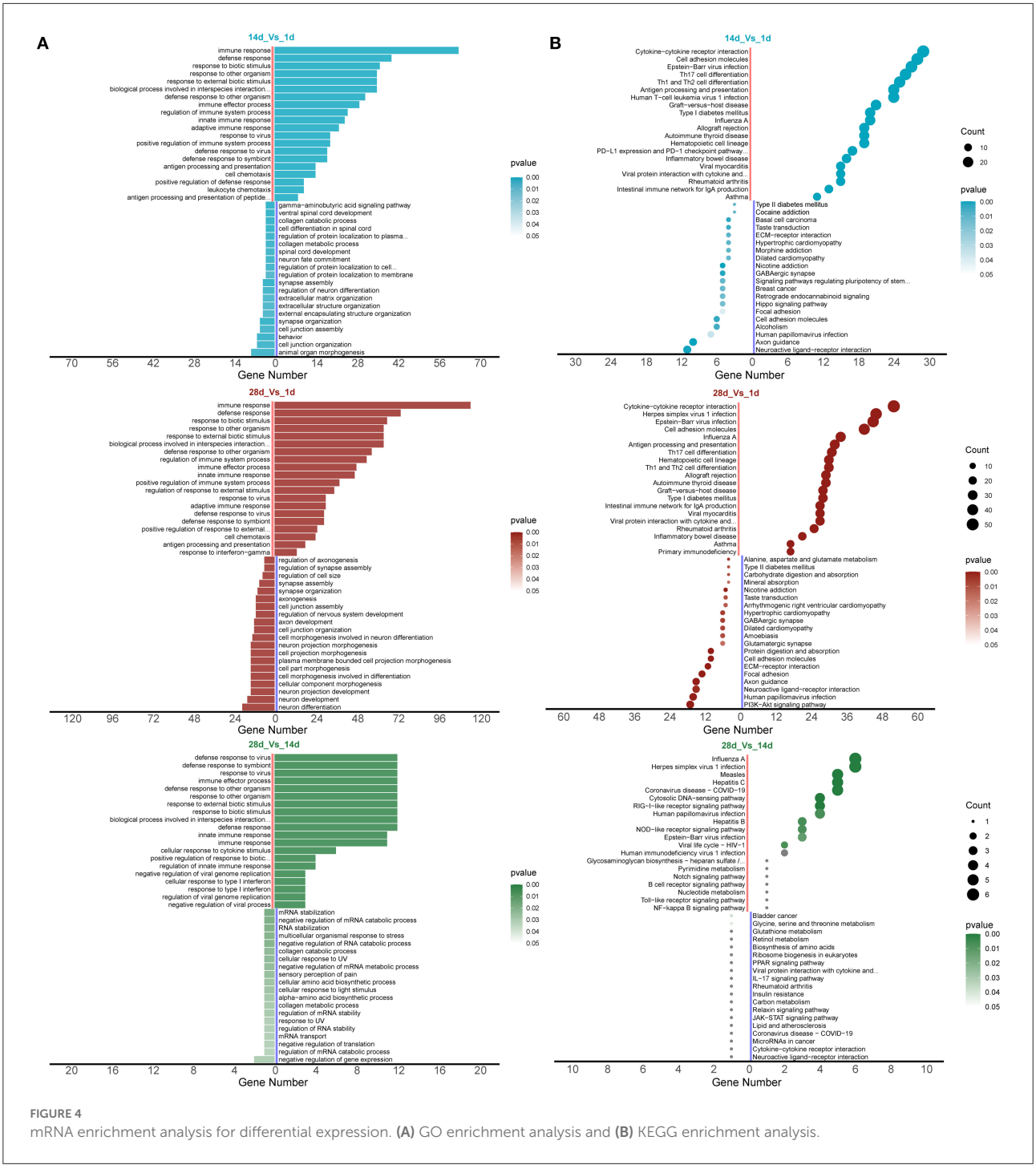


FIGURE 4
mRNA enrichment analysis for differential expression. **(A)** GO enrichment analysis and **(B)** KEGG enrichment analysis.

BHLHE22, STAC3, LOC100517129, OAS1,LOC100519082, IFIT1, and *DHX58*), corresponding to the linkage and enrichment of the innate immune response, immune effector process, and other pathways. *TCONS-00002102* (one lncRNA) is related to seven mRNAs (*IF144, RTP4, BHLHE22, STAC3, LOC100517129, LOC100519082,* and *DHX58*) and is enriched in the innate immune response, immune effector process, and other pathways. A heat map of the topological overlap of interacting mRNAs and lncRNAs in the mRNA–lncRNA–pathway co-expression network in three age groups was created with different color markers; red represents positive correlations, while blue represents negative correlations (Figure 6B). We predicted miRNAs common to mRNA and lncRNA to construct the mRNA–miRNA–lncRNA ceRNA

10.3389/fvets.2022.1031786

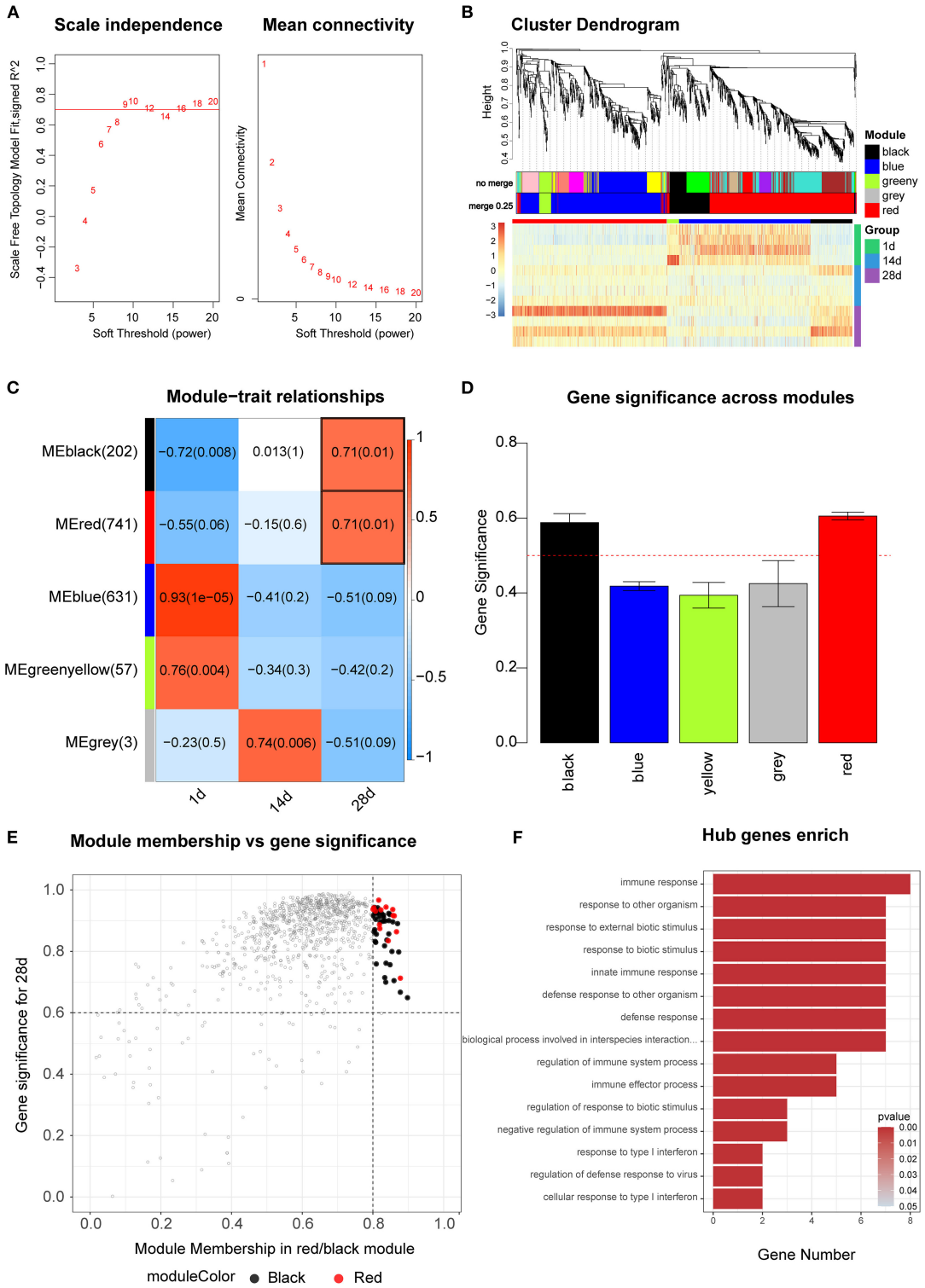


FIGURE 5
WGCNA analysis. **(A)** Scale independence and mean connectivity analysis for various soft threshold powers. **(B)** Clustering dendrograms of mRNAs. Different colors indicate different co-expression modules. Heat map showing the expression profile of protein-coding genes. **(C)** Module–trait relationship. Each row represents a module eigengene, and each column represents a trait. Each cell includes the corresponding correlation and *p*-value. **(D)** Gene significance analysis for each of the 28 d modules. **(E)** Scatter plot of red and black modules. **(F)** GO enrichment analysis of the pathways of the red and black modules.

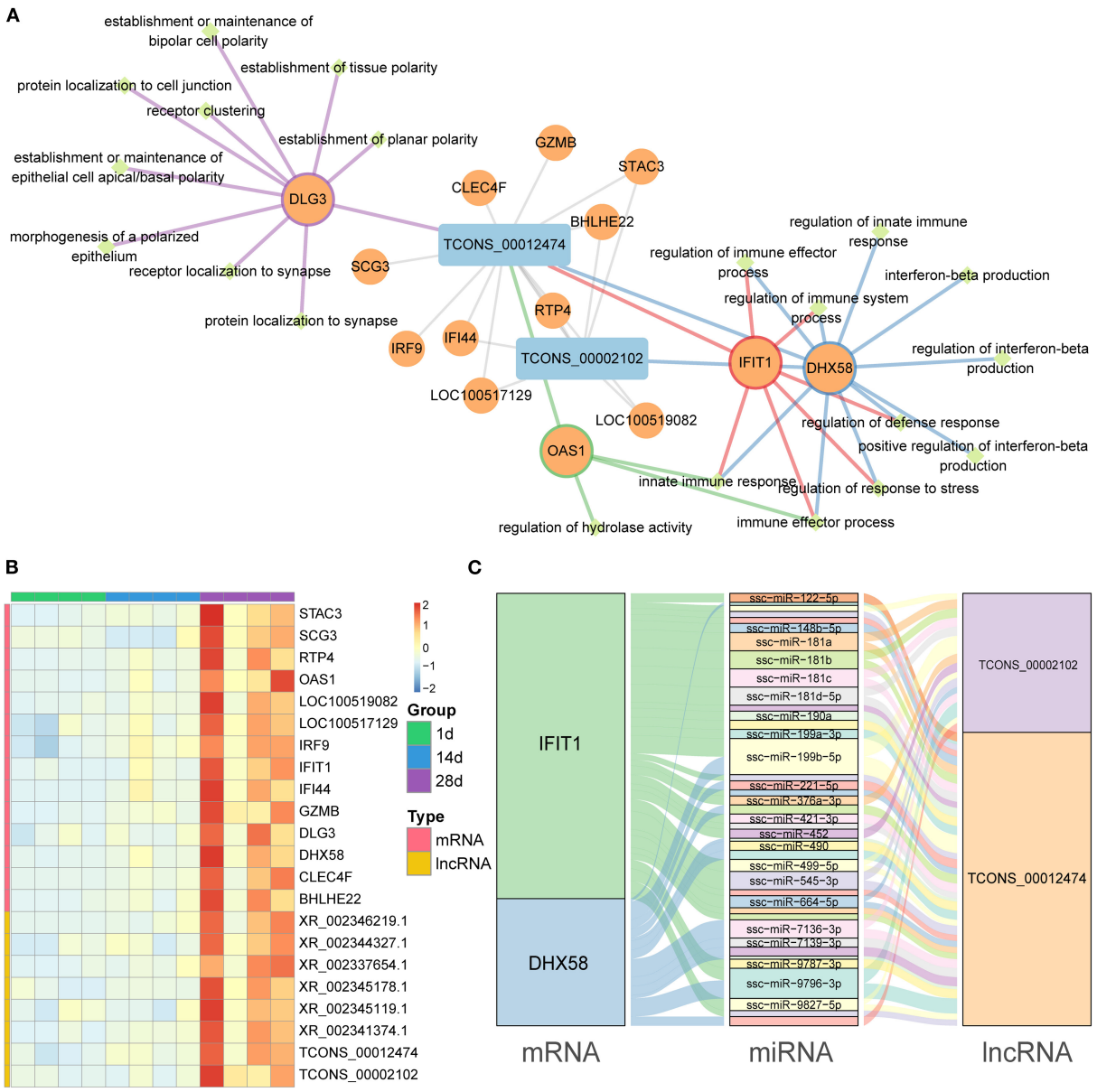


FIGURE 6
mRNA–lncRNA co-expression networks and ceRNA networks. **(A)** Co-expression mRNA–lncRNA network in the red and black modules. The circular nodes represent the mRNAs, and the triangle nodes represent lncRNAs. Gray edges represent mRNA–lncRNA interactions and the other edges represent the mRNA pathways. **(B)** Heat map showing the correlated expression of mRNA and lncRNA in spleen tissue at three different ages. **(C)** ceRNA network shows the relationship between mRNA, miRNA, and lncRNA.

network (Figure 6C, Supplementary Table 11), which included 2 lncRNAs, 2 mRNAs, and 41 miRNAs.

RT-qPCR quantification of mRNAs and lncRNAs

In order to confirm the reliability of RNA-seq data in Meishan piglets, the mRNA (*IFIT1* and *DHX58*) and lncRNA (*TCONS-00002102* and *TCONS-00012474*) genes were selected for RT-qPCR. It can be seen that the RT-qPCR results for these mRNAs and lncRNAs are similar to the RNA-seq results, indicating the accuracy of the RNA-seq data (Figure 7).

Discussion

The spleen is one of the pigs' most important immune organs. Additionally, studies have reported many epidemics

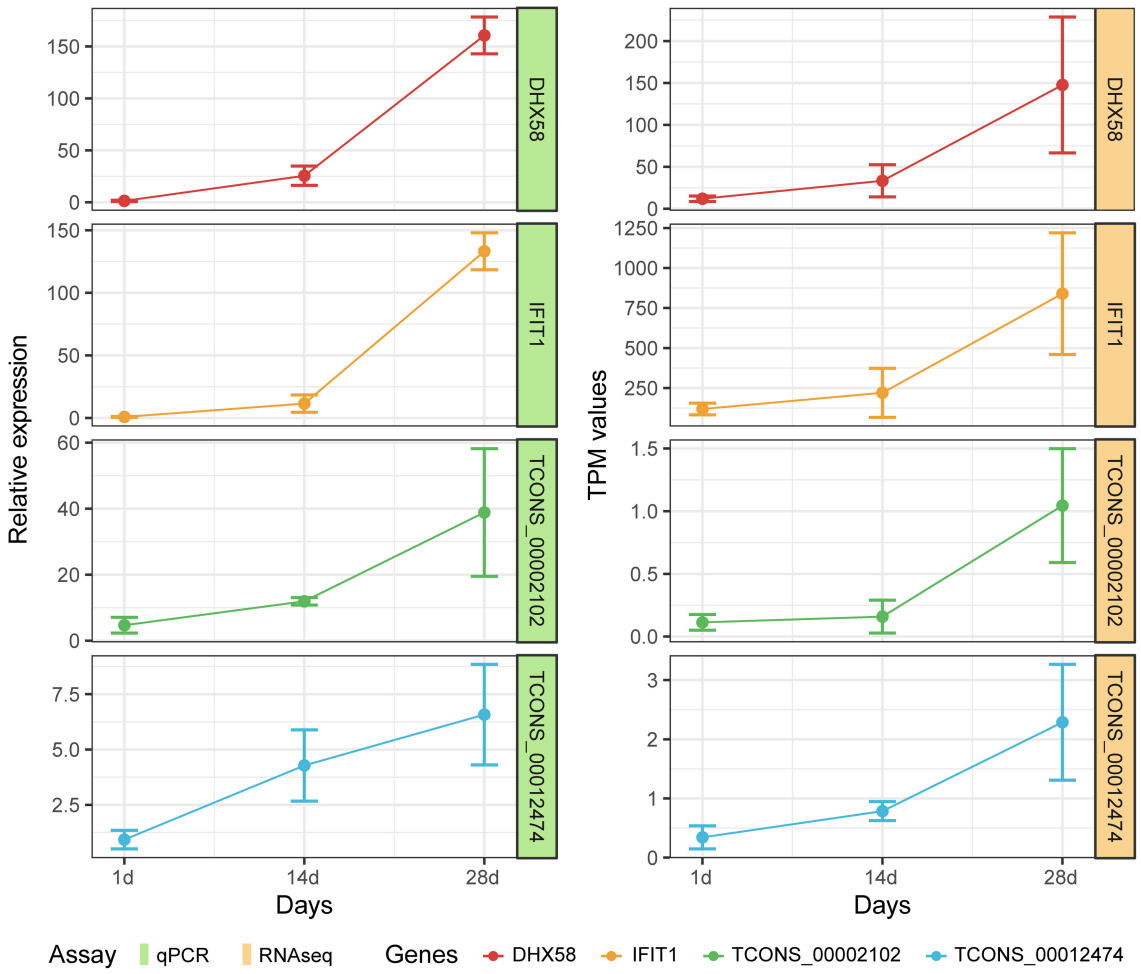


FIGURE 7
Expression patterns of DHX58, IFIT1, TCONS-00002102, and TCONS-00012474 compared with expression patterns obtained by RNA-seq.

in pigs to be associated with spleen damage, such as PRRS, Porcine Circovirus, pseudorabies, and swine fever (43). There are two existing studies on lncRNAs in the spleen of Chinese endemic pigs, and these studies focused on the multiple developmental stages of the pig spleen (22, 43). In our study, Meishan piglets were selected as a model to analyse the transcriptional expression of lncRNA and mRNA in the spleen tissue of piglets for the first time. The differentially expressed 1,806 mRNAs and 319 lncRNAs were identified based on transcriptome expression profiles.

The colostrum is considered to be the first immunization for newborns. Therefore, the three different sample groups selected in this study exhibited large immunological differences. Compared to 14 vs. 1 d and 28 vs. 1 d, the number of differential mRNAs and lncRNAs in 28 vs. 14 d is significantly lower. We speculate that the immune system of piglets starts to establish itself before 14 days, which is in general agreement with the view that the immune system of pigs is established at seven days of life (43).

The results of GO enrichment analysis showed an increase in the expression of genes in pathways related to the immune defense response to viruses. The changes reflect the immunologic function of the spleen (44). In addition, the KEGG results were enriched for upregulating the haematopoietic cell lineage pathway. This result suggests that the spleen may be involved in haematopoietic processes during development (44).

The initial innate immune responses are the first line of defense against viral vectors and help modulate subsequent adaptive immune responses (45). The gene enrichment in the innate immune response and immune effector process pathways discovered by WGCNA analysis demonstrates that the innate immune function of the spleen has been activated. Interferon beta is an important type I interferon that plays an important role in intrinsic antiviral immunity (46), and has anti-tumor, anti-proliferative, and immunomodulatory

functions (47, 48). The enrichment of the interferon beta production pathway indicates that the spleen of piglets already has immunomodulatory functions at 28 days of age.

IFIT1 and *DHX58* were identified as the central genes by mRNA–lncRNA co-expression network analysis. The target genes of the two lncRNAs (*TCONS-00002102* and *TCONS-00012474*) were shown to be *IFIT1* and *DHX58* based on the predicted ceRNA network map. All genes were significantly upregulated with age. *IFIT1*, also known as *ISG56*, is a member of the family of interferon-inducible proteins with tetrapeptide repeats (IFITs) (49). IFITs are important viral restriction factors that have been shown to directly inhibit viral protein synthesis and regulate innate immune signaling pathways. Recently, it was shown that the knockdown of the STAT1 gene, a gene that inhibits Porcine Delta Coronavirus, resulted in a significant increase in PDCoV production, and then downstream interferon-stimulated gene expression was detected, in which IFIT1 expression was found to be substantially decreased (50). This suggests that IFIT1 is important in antiviral replication. Furthermore, Vaishali Sah et al. used IFIT1 as a key indicator of immunity by measuring its expression level after swine fever vaccination (51). It has been reported that interferon can significantly trigger the production of many interferon-induced genes (IFIT1, IFITM3, MX-1, OASL, ISG15, PKR, GBP1, Viperin, BST2, IRF-1, and CXCL10), which play key roles in the resistance to viral infection (52). Bo Yang et al. also showed that the expression of some antiviral and inflammation-related factors was significantly altered after African swine fever virus infection, including the interferon-inducible protein IFIT1 (53). This suggests that IFIT1 plays an important immune function in the host as an interferon-inducible gene. According to our results, IFIT1 also plays a key role in the establishment of splenic immunity in Meishan piglets. *DHX58* is a member of the retinoic-acid-inducible gene (RIG)-like receptor (RLRs) family, which are pattern recognition receptors (PRRs) that trigger an innate immune response against viral infections (54). Previous studies have shown that the mRNA and protein levels of *DHX58* are significantly upregulated in M1 macrophages (55). Moreover, the gene encoding this protein can stimulate macrophages to generate signals that incite the mitochondria to produce inflammasomes, producing inflammatory proteins that play a role in the defense response (56). In addition, it has been shown that *DHX58* negatively regulates the RIG-1 signaling pathway through the competitive binding of viral RNA molecules to RIG-I/MDA5 and inhibits the transcription of type I IFN induced by viral infection (57). Li et al. found that an SNP in the *DHX58* gene was significantly associated with blood parameters in pigs (58). With these results, we speculate that *DHX58* may be one of the crucial genes associated with the immune response in pigs.

A shortcoming of this study is that only mRNA and lncRNA transcripts were analyzed in the spleen of Meishan piglets, and the immune effects of the selected lncRNA transcripts were not explored in depth. This represents a direction for future work.

Conclusion

In summary, the ceRNA networks were constructed by predicting miRNA-targeted lncRNAs and mRNAs, and were screened for the core genes (*DHX58, IFIT1*) and key lncRNAs (*TCONS-00002102, TCONS-00012474*). Additionally, they play a key role in immune defense, the inflammatory response, and other processes. The results of this study contribute to our understanding of the immune function of the spleen in Meishan piglets, lay the foundation for the study of lncRNAs in Meishan pigs, and provide new insights into the function of lncRNAs in spleen tissue. However, this study still has limitations, and more experiments are needed to explore the biological functions of key lncRNAs in order to improve disease resistance in piglets.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The animal study was reviewed and approved by Institutional Animal Care and Use Committee (IACUC) of Yangzhou University (Pig: SYXK(Su)2012-0029).

Author contributions

JS, CX, and WB designed the experiments. JS and CX collected the experimental tissues, analyzed the data, and interpreted the results. JS wrote the manuscript with input from all the authors. SW and ZW participated in designing the structure of the article. All authors have read and approved the final manuscript.

Funding

This work was supported by grants from the Key Research and Development Project (Modern Agriculture) of Jiangsu Province (BE2019341), the Open Competition Mechanism to Select the Best Candidates for the Foundation for Breeding Industry Prosperity of Jiangsu Province, China

(JBGS[2021]098), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Acknowledgments

We thank Oebiotech Corporation (Shanghai, China) for Illumina sequencing.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.1031786/full#supplementary-material

References

1. Raza SHA, Khan R, Cheng G, Long F, Bing S, Easa AA, et al. RNA-Seq reveals the potential molecular mechanisms of bovine KLF6 gene in the regulation of adipogenesis. *Int J Biol Macromol.* (2022) 195:198–206. doi: 10.1016/j.ijbiomac.2021.11.202

2. Gong Y, He J, Li B, Xiao Y, Zeng Q, Xu K, et al. Integrated analysis of lncRNA and mRNA in subcutaneous adipose tissue of ningxiang pig. *Biology.* (2021) 10:726. doi: 10.3390/biology10080726

3. Hombach S, Kretz M. Non-coding RNAs: classification, biology and functioning. *Adv Exp Med Biol.* (2016) 937:3–17. doi: 10.1007/978-3-319-42059-2_1

4. Fathizadeh H, Hayat SMG, Dao S, Ganbarov K, Tanomand A, Asgharzadeh M, et al. Long non-coding RNA molecules in tuberculosis. *Int J Biol Macromol.* (2020) 156:340–6. doi: 10.1016/j.ijbiomac.2020.04.030

5. Bridges MC, Daulagala AC, Kourtidis A. LNCcation: lncRNA localization and function. *J Cell Biol.* (2021) 220:e202009045. doi: 10.1083/jcb.202009045

6. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* (2018) 172:393–407. doi: 10.1016/j.cell.2018.01.011

7. Li Y, Egranov SD, Yang L, Lin C. Molecular mechanisms of long noncoding RNAs-mediated cancer metastasis. *Genes Chromosomes Cancer.* (2019) 58:200–7. doi: 10.1002/gcc.22691

8. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol.* (2018) 19:143–57. doi: 10.1038/nrm.2017.104

9. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* (2021) 22:96–118. doi: 10.1038/s41580-020-00315-9

10. Wu J, Peng X, Qiao M, Zhao H, Li M, Liu G, et al. Genome-wide analysis of long noncoding RNA and mRNA profiles in PRRSV-infected porcine alveolar macrophages. *Genomics.* (2020) 112:1879–88. doi: 10.1016/j.ygeno.2019.10.024

11. Chen J, Zhang C, Zhang N, Liu G. Porcine endemic diarrhea virus infection regulates long noncoding RNA expression. *Virology.* (2019) 527:89–97. doi: 10.1016/j.virol.2018.11.007

12. Fang M, Yang Y, Wang N, Wang A, He Y, Wang J, et al. Genome-wide analysis of long non-coding RNA expression profile in porcine circovirus 2-infected intestinal porcine epithelial cell line by RNA sequencing. *PeerJ.* (2019) 7:e6577. doi: 10.7717/peerj.6577

13. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet.* (2015) 47:217–25. doi: 10.1038/ng.3199

14. Halbur PG, Rothschild MF, Thacker BJ, Meng XJ, Paul PS, Bruna JD. Differences in susceptibility of Duroc, Hampshire, and Meishan pigs to infection with a high virulence strain (VR2385) of porcine reproductive and respiratory syndrome virus (PRRSV). *J Anim Breed Genet.* (1998) 115:181–9. doi: 10.1111/j.1439-0388.1998.tb00341.x

15. Reiner G, Eckert J, Peischl T, Bochert S, Jäkel T, Mackenstedt U, et al. Variation in clinical and parasitological traits in Pietrain and Meishan pigs infected with Sarcocystis miescheriana. *Vet Parasitol.* (2002) 106:99–113. doi: 10.1016/S0304-4017(02)00041-9

16. Chen J, Qi S, Guo R, Yu B, Chen D. Different messenger RNA expression for the antimicrobial peptides beta-defensins between Meishan and crossbred pigs. *Mol Biol Rep.* (2010) 37:1633–9. doi: 10.1007/s11033-009-9576-5

17. Dong L, Li HM, Wang SN, Wang TL Yu LH, Wang HR. Meishan neonatal piglets tend to have higher intestinal barrier function than crossbred neonatal piglets. *Animal.* (2021) 15:100037. doi: 10.1016/j.animal.2020.100037

18. Buschmann H, Pawlas S. A study of porcine lymphocyte populations. II. Characterization of porcine lymphocyte populations. *Vet Immunol Immunopathol.* (1980) 1:215–24. doi: 10.1016/0165-2427(80)90023-9

19. Karte C, Platje N, Bullermann J, Beer M, Höper D, Blome S. Re-emergence of porcine epidemic diarrhea virus in a piglet-producing farm in northwestern Germany in 2019. *BMC Vet Res.* (2020) 16:329. doi: 10.1186/s12917-020-02548-4

20. Jung K, Hu H, Saif LJ. Porcine deltacoronavirus infection: Etiology, cell culture for virus isolation and propagation, molecular epidemiology and pathogenesis. *Virus Res.* (2016) 226:50–9. doi: 10.1016/j.virusres.2016.04.009

21. Xia L, Yang Y, Wang J, Jing Y, Yang Q. Impact of TGEV infection on the pig small intestine. *Virol J.* (2018) 15:102. doi: 10.1186/s12985-018-1012-9

22. Che T, Li D, Jin L, Fu Y, Liu Y, Liu P, et al. Long non-coding RNAs and mRNAs profiling during spleen development in pig. *PLoS One.* (2018) 13:e0193552. doi: 10.1371/journal.pone.0193552

23. Petersen B, Kammerer R, Frenzel A, Hassel P, Dau TH, Becker R, et al. Generation and first characterization of TRDC-knockout pigs lacking γδ T cells. *Sci Rep.* (2021) 11:14965. doi: 10.1038/s41598-021-94017-7

24. Sun J, Zhong H, Du L, Li X, Ding Y, Cao H, et al. Gene expression profiles of germ-free and conventional piglets from the same litter. *Sci Rep.* (2018) 8:10745. doi: 10.1038/s41598-018-29093-3

25. Zhang Y, Xue L, Xu H, Liang W, Wu Q, Zhang Q, et al. Global analysis of alternative splicing difference in peripheral immune organs between tongcheng pigs and large white pigs artificially infected with PRRSV *in vivo*. *Biomed Res Int.* (2020) 2020:4045204. doi: 10.1155/2020/4045204

26. de Sena Brandine G, Smith AD. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res.* (2019) 8:1874. doi: 10.12688/f1000research.21142.1

27. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354

28. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput Biol.* (2017) 24:1138–43. doi: 10.1089/cmb.2017.0096

29. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290–5. doi: 10.1038/nbt.3122

30. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* (2012) 7:562–78. doi: 10.1038/nprot.2012.016

31. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. Pfam: clans, web tools and services. *Nucleic Acids Res.* (2006) 34:D247–51. doi: 10.1093/nar/gkj149

32. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* (2007) 35:W345–9. doi: 10.1093/nar/gkm391

33. Li A, Zhang J, Zhou Z. PLEK a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics.* (2014) 15:311. doi: 10.1201/b16589

34. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* (2013) 41:e166. doi: 10.1093/nar/gkt646

35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:550. doi: 10.1186/s13059-014-0550-8

36. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 40: a universal enrichment tool for interpreting omics data. *Innovation*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

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

38. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* (2005) 4:Article17. doi: 10.2202/1544-6115.1128

39. Zhang X, Cui Y, Ding X, Liu S, Han B, Duan X, et al. Analysis of mRNA-lncRNA and mRNA-lncRNA-pathway co-expression networks based on WGCNA in developing pediatric sepsis. *Bioengineered.* (2021) 12:1457–70. doi: 10.1080/21655979.2021.1908029

40. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell.* (2011) 146:353–8. doi: 10.1016/j.cell.2011.07.014

41. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. *Genome Biol.* (2003) 5:R1. doi: 10.1186/gb-2003-5-1-r1

42. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* (2014) 505:344–52. doi: 10.1038/nature12986

43. Li X, Han X, Sun C, Li G, Wang K, Li X, et al. Analysis of mRNA and long non-coding RNA expression profiles in developing yorkshire pig spleens. *Animals.* (2021) 11:2768. doi: 10.3390/ani11102768

44. Lewis SM, Williams A, Eisenbarth SC. Structure and function of the immune system in the spleen. *Sci Immunol.* (2019) 4:eaau6085. doi: 10.1126/sciimmunol.aau6085

45. Dauletbekov DL, Pfromm JK, Fritz AK, Fischer MD. Innate immune response following AAV administration. *Adv Exp Med Biol.* (2019) 1185:165–8. doi: 10.1007/978-3-030-27378-1_27

46. Zhang X, Yang W, Wang X, Zhang X, Tian H, Deng H, et al. Identification of new type I interferon-stimulated genes and investigation of their involvement in IFN-β activation. *Protein Cell.* (2018) 9:799–807. doi: 10.1007/s13238-018-0511-1

47. Sin WX, Li P, Yeong JP, Chin KC. Activation and regulation of interferon-β in immune responses. *Immunol Res.* (2012) 53:25–40. doi: 10.1007/s12026-012-8293-7

48. Sakamoto I, Tezuka K, Fukae K, Ishii K, Taduru K, Maeda M, et al. Chemical synthesis of homogeneous human glycosyl-interferon-β that exhibits potent antitumor activity *in vivo*. *J Am Chem Soc.* (2012) 134:5428–31. doi: 10.1021/ja2109079

49. Feng B, Zhang Q, Wang J, Dong H, Mu X, Hu G, et al. IFIT1 expression patterns induced by H9N2 virus and inactivated viral particle in human umbilical vein endothelial cells and bronchus epithelial cells. *Mol Cells.* (2018) 41:271–81. doi: 10.14348/molcells.2018.2091

50. Qu H, Wen Y, Hu J, Xiao D, Li S, Zhang L, et al. Study of the inhibitory effect of STAT1 on PDCoV infection. *Vet Microbiol.* (2022) 266:109333. doi: 10.1016/j.vetmic.2022.109333

51. Sah V, Kumar A, Dhar P, Upmanyu V, Tiwari AK, Wani SA, et al. Signature of genome wide gene expression in classical swine fever virus infected macrophages and PBMCs of indigenous vis-a-vis crossbred pigs. *Gene.* (2020) 731:144356. doi: 10.1016/j.gene.2020.144356

52. Fan W, Jiao P, Zhang H, Chen T, Zhou X, Qi Y, et al. Inhibition of African swine fever virus replication by porcine type I and type II interferons. *Front Microbiol.* (2020) 11:1203. doi: 10.3389/fmicb.2020.01203

53. Yang B, Shen C, Zhang D, Zhang T, Shi X, Yang J, et al. Mechanism of interaction between virus and host is inferred from the changes of gene expression in macrophages infected with African swine fever virus CN/GS/2018 strain. *Virol J.* (2021) 18:170. doi: 10.1186/s12985-021-01637-6

54. Dixit E, Kagan JC. Intracellular pathogen detection by RIG-I-like receptors. *Adv Immunol.* (2013) 117:99–125. doi: 10.1016/B978-0-12-410524-9.00004-9

55. Zhong Z, Liang S, Sanchez-Lopez E, He F, Shalapour S, Lin XJ, et al. New mitochondrial DNA synthesis enables NLRP3 inflammasome activation. *Nature.* (2018) 560:198–203. doi: 10.1038/s41586-018-0372-z

56. Murphy MP. Newly made mitochondrial DNA drives inflammation. *Nature.* (2018) 560:176–7. doi: 10.1038/d41586-018-05764-z

57. Vitour D, Meurs EF. Regulation of interferon production by RIG-I and LGP2: a lesson in self-control. *Sci STKE.* (2007) 2007:pe20. doi: 10.1126/stke.3842007pe20

58. Li XY, Han CM, Wang Y, Liu HZ, Wu ZF, Gao QH, et al. Expression patterns and association analysis of the porcine DHX58 gene. *Anim Genet.* (2010) 41:537–40. doi: 10.1111/j.1365-2052.2010.02027.x

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fgene.2022.948240

OPEN ACCESS

EDITED BY
Nuno Carolino,
Instituto Nacional Investigaciao Agraria e Veterinaria (INIAV), Portugal

REVIEWED BY
Zitong Li,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Tiago Bresolin,
University of Illinois at Urbana-Champaign, United States

*CORRESPONDENCE
Gota Morota,
morota@vt.edu
Luiz Lehmann Coutinho,
llcoutinho@usp.br

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

RECEIVED 19 May 2022
ACCEPTED 06 October 2022
PUBLISHED 21 October 2022

CITATION
Novais FJd, Yu H, Cesar ASM, Momen M, Poleti MD, Petry B, Mourão GB, Regitano LCdA, Morota G and Coutinho LL (2022), Multi-omic data integration for the study of production, carcass, and meat quality traits in Nellore cattle.
Front. Genet. 13:948240.
doi: 10.3389/fgene.2022.948240

Multi-omic data integration for the study of production, carcass, and meat quality traits in Nellore cattle

Francisco José de Novais[1], Haipeng Yu[2], Aline Silva Mello Cesar[3], Mehdi Momen[2], Mirele Daiana Poleti[4], Bruna Petry[1], Gerson Barreto Mourão[1], Luciana Correia de Almeida Regitano[5], Gota Morota[2]* and Luiz Lehmann Coutinho[1]*

[1]Department of Animal Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil, [2]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [3]Department of Agri-Food Industry, Food and Nutrition, University of São Paulo, Piracicaba, Brazil, [4]Department of Veterinary Medicine, School of Animal Science and Food Engineering, University of Sao Paulo, Pirassununga, Brazil, [5]Embrapa Pecuária Sudeste, São Carlos, Brazil

Data integration using hierarchical analysis based on the central dogma or common pathway enrichment analysis may not reveal non-obvious relationships among omic data. Here, we applied factor analysis (FA) and Bayesian network (BN) modeling to integrate different omic data and complex traits by latent variables (production, carcass, and meat quality traits). A total of 14 latent variables were identified: five for phenotype, three for miRNA, four for protein, and two for mRNA data. Pearson correlation coefficients showed negative correlations between latent variables miRNA 1 (mirna1) and miRNA 2 (mirna2) (−0.47), ribeye area (REA) and protein 4 (prot4) (−0.33), REA and protein 2 (prot2) (−0.3), carcass and prot4 (−0.31), carcass and prot2 (−0.28), and backfat thickness (BFT) and miRNA 3 (mirna3) (−0.25). Positive correlations were observed among the four protein factors (0.45–0.83): between meat quality and fat content (0.71), fat content and carcass (0.74), fat content and REA (0.76), and REA and carcass (0.99). BN presented arcs from the carcass, meat quality, prot2, and prot4 latent variables to REA; from meat quality, REA, mirna2, and gene expression mRNA1 to fat content; from protein 1 (prot1) and mirna2 to protein 5 (prot5); and from prot5 and carcass to prot2. The relations of protein latent variables suggest new hypotheses about the impact of these proteins on REA. The network also showed relationships among miRNAs and nebulin proteins. REA seems to be the central node in the network, influencing carcass, prot2, prot4, mRNA1, and meat quality, suggesting that REA is a good indicator of meat quality. The connection among miRNA latent variables, BFT, and fat content relates to the influence of miRNAs on lipid metabolism. The relationship between mirna1 and prot5 composed of isoforms of nebulin needs further investigation. The FA identified latent variables, decreasing the dimensionality and complexity of the data. The BN was capable of generating interrelationships among latent variables from different types of data, allowing the integration of omics and

complex traits and identifying conditional independencies. Our framework based on FA and BN is capable of generating new hypotheses for molecular research, by integrating different types of data and exploring non-obvious relationships.

KEYWORDS

Bayesian network, factor analysis, meat quality, latent variables, omics data

Introduction

Meat quality traits, which include meat tenderness, are an important aspect for consumers and are related to the customer's acceptability and beef repurchase (Rust et al., 2008). Meat quality traits are complex and influenced by diet, pre- and post-slaughter management, meat processing, storage methods, genetic factors, and genotype-by-environment interaction (Wheeler et al., 2005; Adzitey, 2011; Guerrero et al., 2013; Njisane and Muchenje, 2016). Most of the biological mechanisms involved in meat quality traits are not completely understood. In this context, systems biology has been proposed to elucidate the flux of molecular information, generating a holistic point of view for complex traits (Ideker et al., 2001). Data from the genome, transcriptome, proteome, microRNAome, and metabolome have been used independently to study the molecular architecture of complex traits and identify important genes, pathways, and networks that underlie economic livestock traits in the last decade (Tizioto et al., 2013; Carvalho et al., 2014; Cesar et al., 2014, 2016; Novais et al., 2019). However, studies using single omic data disregard the interactions among different levels of biomolecules, postulated by the central dogma of molecular biology (Ritchie M. D. et al., 2015).

Complex traits are regulated at different molecular levels, and considerable effort has been made to generate multi-level studies, integrating different omic data to understand the inherent biological meaning of livestock traits (Widmann et al., 2013; Suravajhala et al., 2016). However, omic data integration using a hierarchical analysis approach or considering just the common pathway enrichment may not reveal non-obvious relationships that exist among omic data (Misra et al., 2018). In this context, efforts to develop approaches to data omic integration have been proposed (Huang et al., 2017).

Factor analysis (FA) reduces the dimensionality of data, inferring latent (hidden) variables to explain dependencies among observed variables that share common variations (Meng et al., 2016). Furthermore, the Bayesian network (BN) has the potential to generate relationships among phenotypes and molecules by a graph-based model of joint multivariate probability distributions that represent conditional independence between variables (Rodin and Boerwinkle, 2005). Here, phenotypes of production, carcass, meat quality, and multi-omic data were fitted into the FA and BN framework to explore the potential biological interrelationships to generate new hypotheses for complex traits in beef cattle.

Materials and methods

Animals and phenotypes

A total of 386 Nellore steers born between 2009 and 2011 at the Brazilian Agricultural Research Corporation (EMBRAPA/Brazil) were initially included in this study. The animals were raised in feedlots under identical diets, and environmental conditions, and slaughtered at age of 25 months. More details regarding animals, diet, and experimental design can be found in Cesar et al. (2014). The animals were handled and managed according to the Institutional Animal Care and Use Committee Guidelines from the Brazilian Agricultural Research Corporation—EMBRAPA approved by the president, Dr. Rui Machado.

Carcass ultrasound evaluations were performed by trained field technicians and followed the standards set by the Ultrasound Guidelines Council (UGC; www.ultrasoundbeef.com). An Aquila Pie Medical (Pie Medical Inc., Maastricht, Netherlands) equipped with a 172 mm-long linear transducer with a frequency of 3.5 MHz was used to measure the initial ribeye area (REAi) and initial backfat thickness (BFTi) obtaining sectional images of the longissimus dorsi (LD) muscle between the 12th and 13th ribs. The images were stored and measurements were obtained by ODT Eview R (Pie Medical Inc., Maastricht, Netherlands).

The details of carcass and meat quality trait evaluations were previously described by Nascimento et al. (2016). The visceral organs were removed during slaughter, and the heart, kidney, liver, and perirenal, pelvic, and inguinal fats were weighed. Carcasses were weighed and chilled for 24 h at 5°C. The carcass was weighted at 24 h, and the carcass depth was measured on the fifth rib from top to bottom, measuring the distance from the sternum to the middle of the spine where the marrowbone passes.

Steaks of 2.54 cm thick from the LD muscle between the 12th and 13th ribs were collected 24 h after slaughter. Steaks were vacuum packed and used to measure the shear force (SF; Kg), backfat thickness (BFT; mm), ribeye area (REA; cm^2), myofibrillar fragmentation index (MFI), color parameters

(L* = lightness, a* = redness, and b* = yellowness), intramuscular fat (IMF; percentage), pH at 24 h, moisture, water holding capacity, and cook loss. Briefly, the final backfat thickness (BFTf) was measured using a ruler in millimeters (Filho, 2000). Color parameters L*, a*, and b* were measured after exposing the steaks to atmospheric oxygen for 30 min prior to analysis using a Hunter Lab colorimeter model MiniScan XE with Universal Software v. 4.10 (Hunter Associates Laboratory, Reston, VA), illuminant D65, and 10° standard observer. Additionally, muscle pH was measured at three locations across the steak using a Testo pH measuring instrument model 230 (Testo, Lenzkirch, Germany). The final ribeye area (REAf) was calculated as the area of LD muscle using a grid. Cooking losses were measured as the weight difference between the steaks before and after cooking. For IMF, approximately 100 g of muscle samples, previously lyophilized and ground, were obtained using an Ankom XT20 extractor as described in AOCS official procedure Am 5-04 (Horwitz, 2000). The myofibrillar fragmentation index was determined according to Hopkins et al. (2000). The SF values were obtained from 2.54 cm thick steaks after 24 h of aging at 2°C in a cold chamber using the texture analyzer TA-XT2i coupled to a Warner–Bratzler blade with 1.016 mm thickness.

mRNA data processing and WGCNA

For total RNA extraction, a sample of 100 mg of the LD muscle was processed using the Trizol reagent (Life Technologies, Carlsbad, CA, United States), following the manufacturer's guidelines. After extraction, RNA integrity was verified using the Bioanalyzer 2100 (Agilent, Santa Clara, CA, United States), and the samples presenting RNA integrity numbers lower than 7.0 were removed from further analysis. A total of 2 µg of RNA from each sample was used for the cDNA library preparation, in accordance with the protocol described in the TruSeq RNA Sample Preparation kit v2 guide (Illumina, San Diego, CA, United States). The libraries were sequenced using the HiSeq2500 ultra-high-throughput sequencing system (Illumina, San Diego, CA, United States) with the TruSeq SBS kit v3-HS (200 cycles). All sequencing analyses were performed at the ESALQ Genomics Center (Piracicaba, São Paulo, Brazil).

The FastQC software v0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was applied to check the quality of the sequencing data. Low-quality reads were filtered and adapter sequences were trimmed using Seqyclean package version 1.4.13 (Neapolitan, 2004). The details of data acquisition were previously described by de Lima et al. (2020).

The read alignment was carried out against the bovine reference genome *Bos taurus* ARS-UCD1.2 (available at the Ensembl database https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1) and read counts using STAR software (Spliced Transcripts Alignment to a Reference) version 2.7 (Dobin and Gingeras, 2015) with the Ensembl (release 95, January 2019) gene annotation file. Subsequently, genes with zero counts for all samples were removed. Next, the genes were filtered by the counts different from zero in at least 70% of the samples and counts per million (CPM) > 5 using the EdgeR Bioconductor package (Chen H.-J. et al., 2018). This was followed by normalizing counts using the DESEq2 Bioconductor package (Love et al., 2014), and a batch effect was identified using the limma R package (Ritchie M. E. et al., 2015).

Clustering analysis was performed on the mRNA dataset using the weighted gene co-expression network analysis (WGCNA) R package (Langfelder and Horvath, 2008). To measure the connectivity among genes, an adjacency matrix was generated by calculating the Pearson's correlation coefficients among all genes and raising it to a power β (soft threshold) of 6, which is chosen using a scale-free topology criterion (R^2 = 0.8). Modules containing at least 30 genes were retained. Modules with hub genes that had a module membership (MM) > 0.95 and gene significance (GS) with a p-value < 0.001 were kept for further analysis. Enrichment analysis was performed using MetaCore software (MetaCore, 2021) to elucidate biological processes and pathways represented by the hub genes of modules.

miRNA and data acquisition

Small RNA libraries were constructed from 1 µg of total RNA from each sample using the Illumina TruSeq small RNA Sample Prep Kit (Illumina Inc, San Diego, CA, United States), in accordance with the manufacturer's protocol. High Sensitivity DNA Chip and an Agilent 2100 Bioanalyzer (Agilent Technologies) was used to determine library quality and qPCR with the KAPA Library Quantification kit (KAPA Biosystems, Foster City, CA, United States) for quantification. Sequencing was performed using a Miseq Reagent Kit v3 for 150 cycles in an Illumina Miseq Sequencing System (Illumina Inc., San Diego, CA, United States). The Illumina CASAVA v1.8 was used to generate and de-multiplex the raw fastq sequences. The quality of Illumina deep sequencing data was determined using the FastQC program (version 0.9.5) (Andrews, 2010). Adapters and low-quality reads were trimmed using Cutadapt (version 1.2.1) (Martin, 2011). Filtered reads were then processed following the mirDeep2 analysis pipeline (Friedländer et al., 2012). Sequences were aligned to the *Bos taurus* ARS-UCD.1.2 reference genome (available at the Ensembl database (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1). Only alignments with zero mismatches in the seed region (first 18 nucleotides of a read sequence) of a read mapped to the genome were retained. More details about data acquisition were provided by Kappeler et al. (2019).

Briefly, miRNAs with zero counts for all samples were removed. Next, the miRNAs were filtered by the counts that are different from zero in at least 70% of the samples and CPM >5 using the EdgeR Bioconductor package (Chen Y.

et al., 2018). The miRNA counts were normalized using the DESEq2 Bioconductor package (Love et al., 2014), and the limma R package was used to identify a batch effect (Ritchie M. E. et al., 2015).

Proteome and data acquisition

The details for data acquisition and processing are previously described in Poleti et al. (2018). Frozen muscles (500 μg) of 106 animals were ground on liquid nitrogen, then transferred to a microcentrifuge tube, and weighed to minimize protein degradation. The muscle was homogenized in 2.5 ml lysis buffer containing 8 M urea , 2 M thiourea, 1% DTT, 2% CHAPS, and 1% protease inhibitor cocktails (Sigma-Aldrich) in an ULTRA-TURRAX® IKA homogenizer on ice for 2 min. The extracts were vigorously shaken for 30 min on ice and centrifuged at 10,000 x g for 30 min at 4°C. The supernatants were collected, the total protein concentration was determined by the PlusOne 2-D Quant Kit (GE Healthcare), and then stored at −80°C for further analysis.

The protein extract was desalted with a 3-kDa cutoff Amicon® Ultra centrifugal filter (Millipore, Ireland), where the lysis buffer was exchanged using a solution of 50 mm ammonium bicarbonate and 2 M urea five times. The concentration of the retained protein solution was quantified using a Bradford Protein Assay Kit (BioRad). For protein digestion, 50 μg of proteins of each sample were denatured with 25 μL of 0.2% RapiGest SF (Waters Corporation, United States) at 80°C for 15 min, reduced with 2.5 μL of 100 mm dithiothreitol (DTT) (Sigma, United States) at 60°C for 30 min, and alkylated with 2.5 μL of 300 mm iodoacetamide (AA) (Sigma, United States) at room temperature in the dark for 30 min. Enzymatic digestion was performed with sequencing grade modified trypsin (Promega) at a 1:100 (w/w) enzyme: protein ratio at 37°C for 16 h. Digestion was stopped by the addition of 10 μL of 5% (V/V) trifluoroacetic acid and incubated at 37°C for 90 min to hydrolyze the RapiGest (Yu et al., 2003). The peptide mixture solution was then centrifuged at 18,000 x g for 30 min at 6°C. The supernatant was transferred to a new vial, dried down in a vacuum centrifuge, and stored at −20°C.

Qualitative and quantitative bidimensional nanoUPLC tandem nanoESI-HDMSE analyses were conducted using both 1-h reversed-phase gradient from 7% to 40% (v/v) acetonitrile (0.1% v/v formic acid) and 500 nL*min^{-1} on a nanoACQUITY UPLC 2D Technology system (Gilar et al., 2005). A nanoACQUITY UPLC HSS T3 1.8 μm, 75 μm × 15 cm column (pH 3) was used in conjunction with a reverse-phase (RP) XBridge BEH130 C18 5 μm 300 μm × 50 mm nanoflow column (pH 10). The ion mobility cell was activated and filled with nitrogen gas, which operates at the cross-section resolving power of at least 40 Ω/ΔΩ (Lalli et al., 2013). The effective resolution has the conjoined ion mobility of >1.5 M FWHM (Silva et al., 2014). The ionization of samples was performed using a NanoLockSpray ionization source (Waters, Manchester, United Kingdom) in the positive ion mode nanoESI (+). The mass spectrometer was calibrated with an MS/MS spectrum of [Glu1]-fibrinopeptide B human (Glu-Fib) solution (100 fmol*uL^{-1}) delivered through the reference sprayer of the NanoLockSpray source. Data acquisition was performed using a Synapt G2-S HDMS mass spectrometer (Waters, Manchester, United Kingdom). A mass–charge value ranges from m/z 50 to 2000.

Mass spectrometry data were acquired with Waters MassLynx v.4.1 software and processed using Progenesis QI for Proteomics (QIP) 2.0 software (Nonlinear Dynamics, United Kingdom). Progenesis QIP software was used to run alignment, peak picking, ion drift time data collection, ion abundance measurements, normalization, quantification, peptide and protein identification, and statistical analysis. The processing parameters for Progenesis included the following: automatic tolerance for precursor and product ions based on peptide identification and normal distribution (Geromanos et al., 2009), one missed cleavage, carbamidomethylation of cysteine as a fixed modification, and oxidation of methionine as variable modification. For protein identification and quantification, the obtained raw data were searched against a Nellore transcriptome database built from RNA-sequencing data from LD muscle. Data quality assessment was performed accordingly (Souza et al., 2017), and proteins were selected based on the detection and identification in at least 80% of biological samples. The assembled data were compared to the NCBI's UniProt database (https://www.uniprot.org/) as functional analysis.

Factor analysis

This section closely follows the work of Yu et al. (2020) and Momen, et al. (2021). The exploratory factor analysis (EFA) was applied to search the structure of underlying latent variables (factors) that drive the observed phenotypes and omic data. First, the caret R package (Kuhn, 2008) was used to check collinearity, and one of the features with correlation >0.9 was removed. Then, the Kaiser–Meyer–Olkin (KMO) test was applied to measure the sampling adequacy using the psych R package (Revelle, 2017) assessing the factor ability of the data (Cerny and Kaiser, 1977). The measure of sampling adequacy ranges between 0 and 1, and values closer to 1 are preferred. Here, KMO >0.7 was considered acceptable. The number of underlying latent variables q was determined using a parallel analysis (Horn, 1965) using the psych R package, as described in more detail in a previous work of our group (Momen et al., 2021). The EFA model is given as a function of latent factor scores.

$$Y = \Lambda F + \varepsilon,$$

where Y is a $p \times n$ matrix of p molecular features or phenotypes of n animals, Λ is the $p \times q$ matrix of factor loading connecting the relation between features and latent common factors, F is the $q \times n$ matrix of latent factor scores, and ε is the $p \times n$ vector of unique effects that is not explained by q underlying common factors. The variance–covariance matrix of Y is

$$\Sigma = \Lambda\Phi\Lambda' + \Psi,$$

where Σ is the $p \times p$ variance–covariance matrix of phenotypes, Φ is the variance of factor scores, and Ψ is a $p \times p$ diagonal matrix of unique variance. The elements of Λ, Φ, and Ψ are parameters of the model to be estimated from the data. With the assumption of $F \sim \mathcal{N}(0, I)$, Λ and Ψ were estimated by maximizing the log-likelihood of $\mathcal{L}(\Lambda, \Psi|Y)$ using the R package psych (Revelle, 2017) along with a varimax rotation (Kaiser, 1958). A parallel analysis was performed to determine the number of underlying factors. A feature having loading > |0.55| was assigned to only one of the factors based on the factor loadings.

The Bayesian confirmatory factor analysis (BCFA) is an alternative to frequentist CFA generating an important role in the assessment of the reliability and validity of latent variables. We fitted BCFA to estimate the factor scores according to the phenotype-factor structure inferred from the earlier EFA step. BCFA was applied to concatenated data, including phenotypes, proteins, miRNA, and the hub genes of modules obtained from WGCNA. Briefly, the blavaan R package (Merkle and Rosseel, 2018) was used with three Markov Monte Carlo chains, each with 6,000 Gibbs samples after 6,000 burn-in. Then, the posterior means of the factor scores of latent variables were estimated and treated as the new phenotypes for further analysis.

Bayesian network

In the Bayesian network (BN), a direct acyclic graph is generated, and each random variable is associated with a node, the edges represent conditional dependency between variables, whereas the absence of an edge implies that the variables are conditionally independent of other variables (Choi, 2015). The details of BN procedures can be found in more detail in Yu et al. (2019) and Momen et al. (2021). Briefly, the BN structure learning with the bnlearn R package (Scutari, 2010) was applied to study the probabilistic relationships among the omic and latent variables. The BN is given by

$$BN = (G, X_V),$$

where G represents a direct acyclic graph composed of nodes (V) connected by edges (E), describing the probabilistic relationships and the vector $X_V = (X_1, \ldots, X_k)$ where k is the random variable (Yu et al., 2019). The joint probability of distributions is therefore given by

$$P(X_V) = \prod_{v=1}^{k} P(X_V | P_a(X_V)),$$

where $P_a(X_V)$ expresses a set of parent nodes of X_V. The score-based (hill climbing and tabu) and hybrid algorithms (max–min hill climbing and general 2-phase restricted maximization) were used to perform structure learning (Scutari, 2010). Candidate networks were compared based on the Bayesian information criterion (BIC) and Bayesian Gaussian equivalent score (BGe). The BIC score was calculated as a criterion for the selection of the candidate model, and BGe reflects the posterior probability of the networks. A larger BIC score is preferred since it is rescaled by −2 in the bnlearn R package. In addition, 1,000 bootstrapping replicates were used to estimate the uncertainty of the edge's strength and the direction of the network. Edges showing presence in at least 80% (strength) among all the 1,000 models were kept in the BN through model averaging.

Results

Data preprocessing for analysis

In this study, we investigated the effective application of FA and BN framework to generate networks with biological meaning on for three different phenotypic categories: 1) production trait category included pre-feedlot body weight (BWi), post-feedlot body weight (BWf), initial backfat thickness (BFTi), and initial ribeye area (REAi); 2) carcass trait category included final backfat thickness (BFTf), final ribeye area (REAf), hot carcass weight (carcass_hot), cold carcass weight (carcass_cold), carcass depth (carcass_depth), kidney fat content (fat_kidney), and pelvis fat content (fat_pelvis); and 3) meat quality category included the shear force at 24 h (SF), pH at 24 h (pH), meat moisture (moisture), free water (water_free), water-holding capacity (w_ret_cap), cooking weight loss (cook_loss), color parameters (L*, a*, and b*), myofibrillar fragmentation index (MFI), and intramuscular fat (IMF) along with three different omic datasets: 1) mRNA sequencing, 2) miRNA sequencing, and 3) protein abundance.

Pearson's correlations (Figure 1) showed that BWf, carcass_hot, and water_free were highly correlated with carcass cold and water-holding capacity (correlation >0.9), therefore; they were removed for further analysis to avoid duplicate information. For example, the correlation between water_free and w_ret_cap was −1 because both traits represent oppositional and complementary information (Pearce et al., 2011).

For the RNA-Seq (mRNA) data, after the quality control and filtering procedure, 13,023 genes were included in WGCNA. The WGCNA method identified 20 modules, and two modules (mRNA1 and mRNA2) showed module membership (MM) > 0.95 and gene significance p-value <

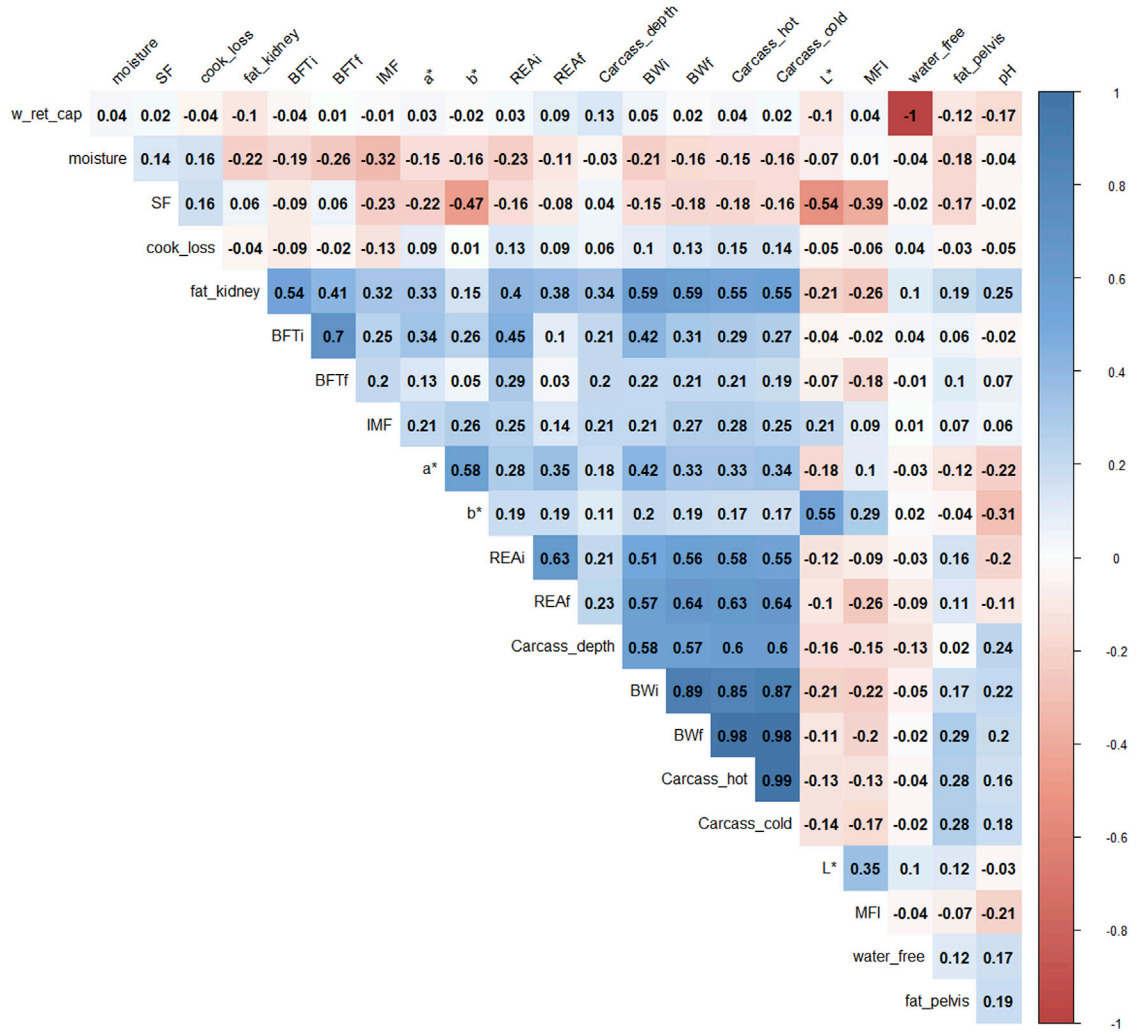


FIGURE 1
Correlation plot of 22 phenotypes. The degree of shading and the value reported correspond to the correlations among the traits. BWi: pre-feedlot body weight; REAi: initial ribeye area by ultrasonography; REAf: final ribeye area on steak; BFTi: initial backfat thickness by ultrasonography; BFTf: final backfat thickness by carcass; fat_pelvis: pelvis fat content at carcass; fat_kidney: kidney fat content; carcass_hot: hot carcass weight; carcass_cold: cold carcass weight; carcass_depth: carcass depth; pH: pH at 24 h; water_free: free water; w_ret_cap: water holding capacity; moisture: meat moisture; SF: shear-force; MFI: miofibrilar fragmentation index; L*, a*, b*: color parameters; and IMF: intramuscular fat.

0.001. The mRNA1 module was composed of seven hub genes and the mRNA2 module of four hub genes (Supplementary Table S1).

A total of 192 miRNAs were used for further analysis after the preprocessing steps. One animal was excluded as an outlier. After normalization, limma was used to identify a batch effect (Figure 2). Principal component analysis (PCA) revealed clusters based on the total counts of samples (Figure 2A). limma was applied to remove the batch effect in the miRNA data for further analysis (Figure 2B).

For proteomic data, 159 proteins from 106 animals were used in the analysis after the quality control steps. PCA was applied and a batch effect due to the equipment used was identified (Figure 3A). The batch effect was accounted for by normalizing every data separately (Figure 3B).

Exploratory and Bayesian confirmatory factor analysis

The factor analyses were performed using a subset of 102 animals that have phenotypes, miRNA, mRNA, and protein data. First, the phenotypes, miRNA, and protein data were used individually to fit an exploratory factor analysis (EFA). EFA can reduce data dimension without any prior assumptions about the observed data and latent factors structures. The parallel analysis suggested that phenotypes,

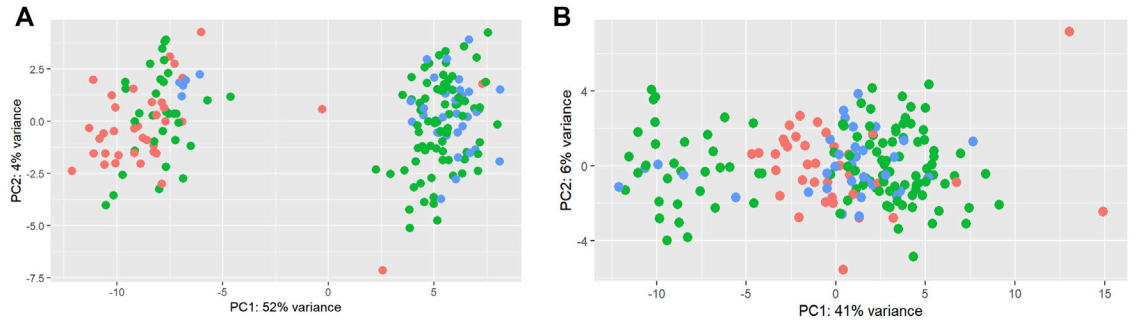


FIGURE 2
Principal component analysis of total counts as a batch effect in miRNAs. Principal component analysis of miRNAs before **(A)** and after **(B)** the limma batch effect normalization. The total counts refer to the total number of reads per sample. Three colors were used to represent 1) samples with a higher total number of reads: higher than mean +standard deviation (345,281 reads) (blue); 2) samples with a lower total number of reads: lower than mean—standard deviation (125,193 reads) (red); 3) samples with the average total number of reads: between mean > + standard deviation and mean < + standard deviation (green).

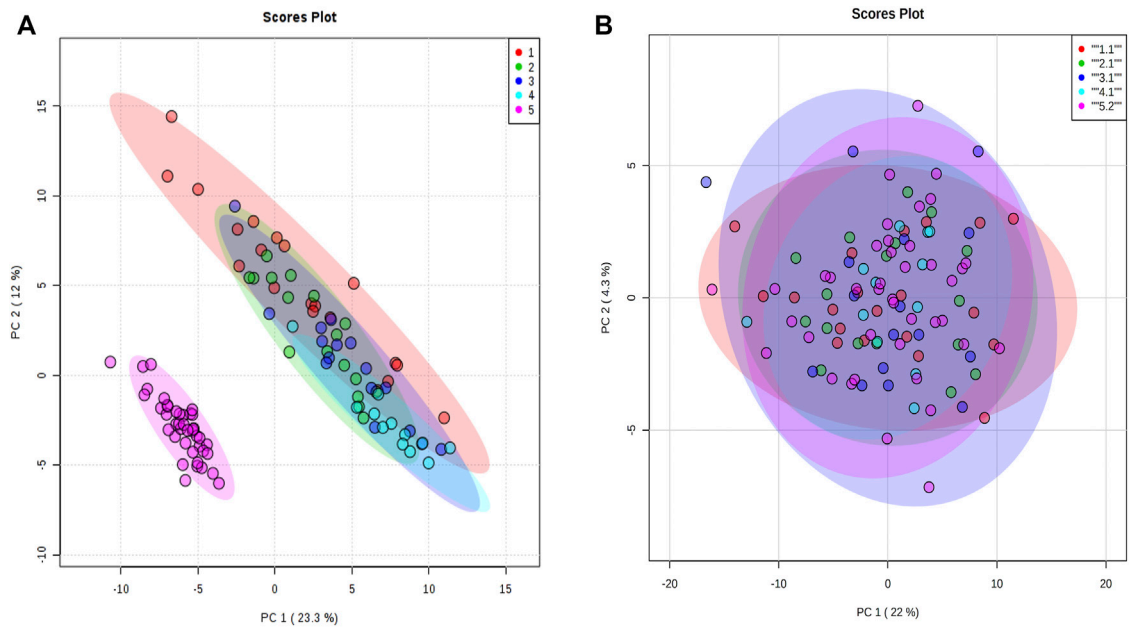


FIGURE 3
Principal component analysis of proteomic data. **(A)** Principal component analysis with all animals normalized together. **(B)** Principal component analysis when animals were normalized separately by equipment acquisition. The colors denote the equipment batch effect. The samples in red, green, dark blue, and blue colors are from equipment 1. The samples in pink are from equipment 2.

miRNA, and protein data were composed of five, ten, and eight latent variables, respectively. Each omic dataset was assigned to a factor according to the highest loading value (>|0.5|), filtering some latent variables composed of a few features. The final underlying latent structures from EFA of the phenotype, miRNA, and protein data are shown in Figures 4, 5.

The BCFA was used to estimate factor loadings and scores based on the structure obtained from the EFA analysis, assuming that these latent variables determine the observed phenotypes and molecular profile levels (Supplementary Tables S2, S3).

The five phenotype latent factors showed strong contributions to the observed phenotypes, with

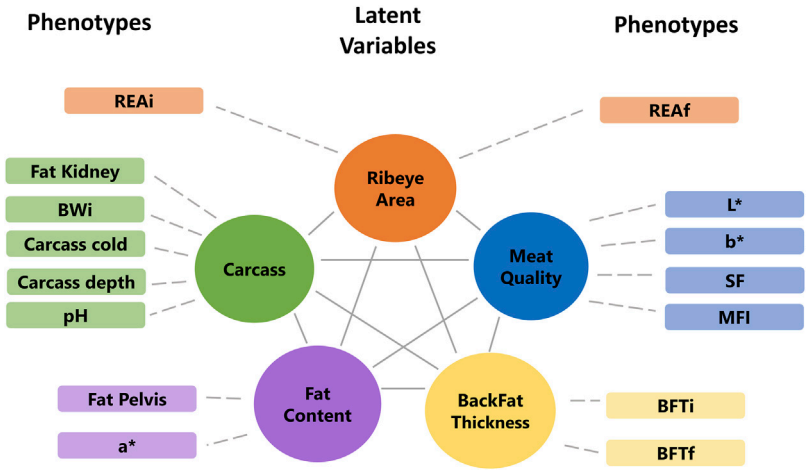


FIGURE 4
Final underlying latent structures of phenotypes generated by exploratory factor analysis. BWi: initial body weight on feedlot trial; REAi: initial ribeye area by ultrasonography; REAf: final ribeye area on steak; BFTi: initial backfat thickness by ultrasonography; BFTf: final backfat thickness by carcass; SF: shear-force; MFI: myofibrillar fragmentation index; and L*, a*, b*: color parameters.

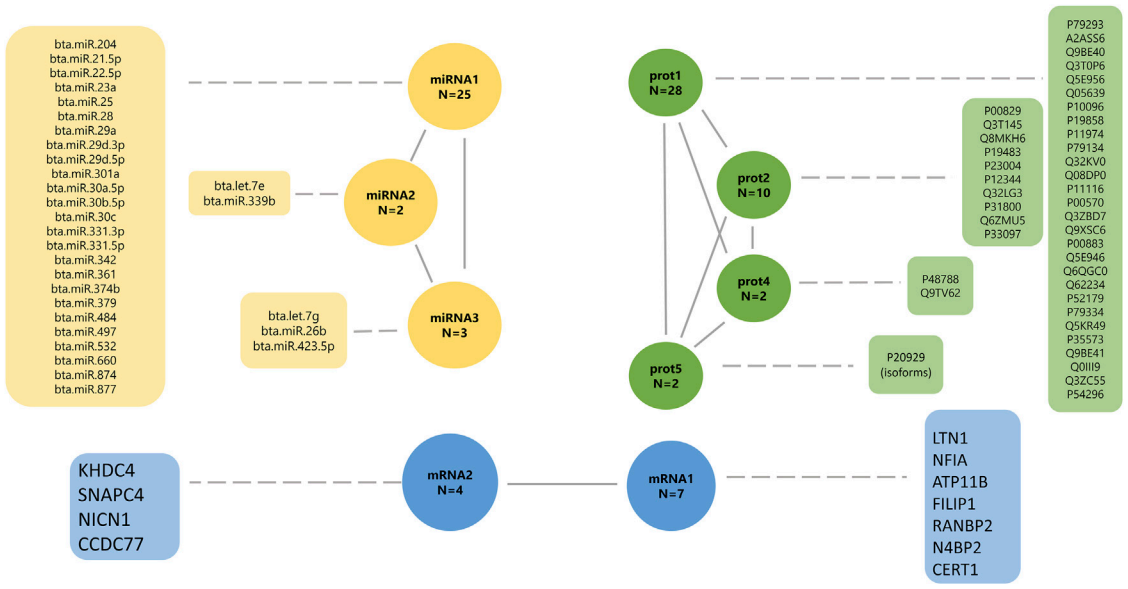


FIGURE 5
Final underlying latent structures of miRNA (yellow), proteins (green), and mRNA (blue) generated by exploratory factor analysis. N denotes the total number of features in each latent variable.

standardized regression coefficients ranging from 0.989 to 0.986 for backfat thickness, −0.993 to 0.956 for meat quality, 0.654 to 1 for the carcass, 0.942 to 0.992 for fat content, and 0.973 to 0.991 for ribeye area. The seven latent variables for miRNA and protein also showed strong contributions to the molecular level profiles, with standardized regression coefficients ranging from -0.999 to 0.999 for factor mirna1 (miRNA), −0.971 to 0.979 for factor mirna2 (miRNA), −0.914 to 0.989 for factor mirna3 (miRNA), 0.842 to 0.990 for factor prot1 (protein), 0.774 to 0.973 for factor prot2 (protein), 0.963 to 0.997 for factor prot4 (protein), and 0.976 to 0.990 for factor prot5 (protein).

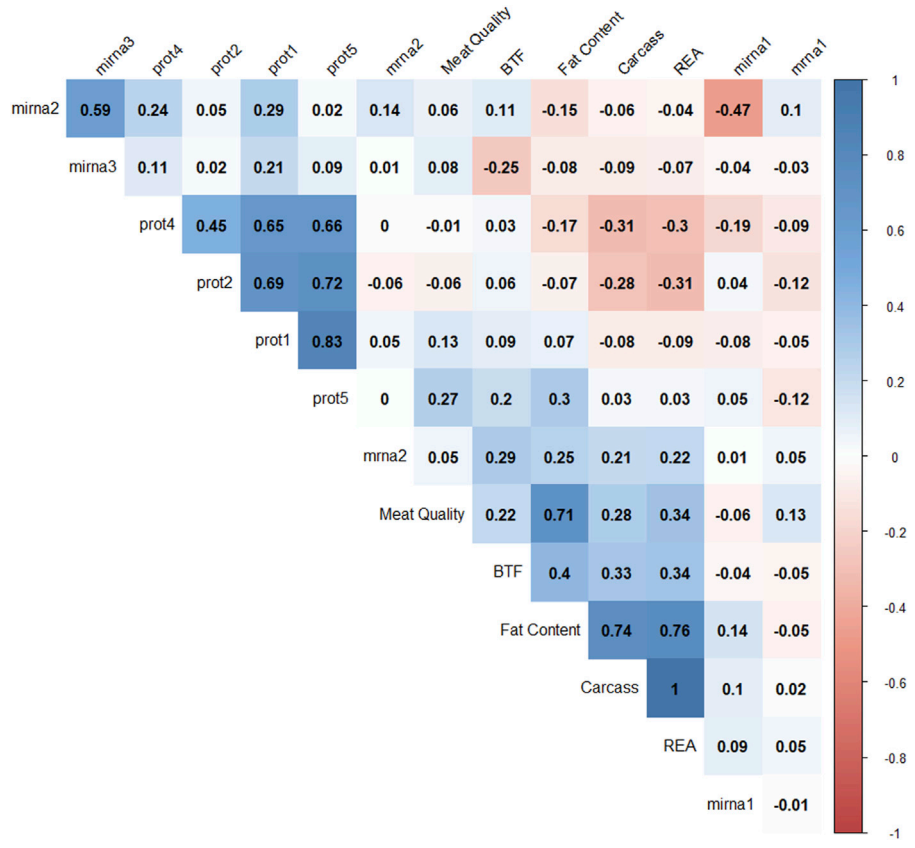


FIGURE 6
Correlation plot of 14 factor scores. The degree of shading and the value reported correspond to the correlation between each pair of latent variables.

The latent factor backfat thickness (BFT) had a positive contribution to BFTi and BFTf (0.989 and 0.986, respectively; Supplementary Table S2), indicating that larger values for the latent factor can be interpreted as a greater thickness on the backfat content. The latent factor meat quality has a positive contribution to shear force (0.956; Supplementary Table S2), and a negative contribution to the colors b*, L*, and MFI (−0.993, −0.959, and −0.785, respectively) indicating that lower values on the latent factor can be interpreted as more tender meat. The latent factor carcass showed the largest positive contributions to traits describing carcass (e.g., weight to carcass cold, 1; weight to carcass depth, 0.990; weight to the kidney's fat content, 0.987; and pH of meat at 24 h, 0.863), suggesting that this latent factor is an overall representation of carcass. The latent factor ribeye area (REA) has a strong positive contribution to the REAf and REAi (0.991 and 0.973, respectively; Supplementary Table S2), indicating that larger values for the latent factor can be interpreted as a greater ribeye area.

The latent factor mirna1 has a positive contribution to 18 miRNAs (0.876–0.999; Supplementary Table S2), and a negative contribution to seven miRNAs (−0.995 to −0.999, respectively). The mirna2 latent variable has a positive contribution to miRNA "bta.let.7e" (0.979; Supplementary Table S2), and a negative contribution to miRNA "bta.miR.339b" (−0.971, respectively; Supplementary Table S2). The latent factor mirna3 has a positive contribution of two miRNAs, "bta.let.7 g" (0.889) and "bta.miR.26b" (0.987), and a negative contribution to miRNA "bta.miR.423.5p" (−0.914). The latent factors prot1, prot2, prot4, and prot5 have a positive contribution to all proteins, including 28 proteins (0.842–0.990), 10 proteins (0.774–0.973), two proteins (0.976–0.997), and two proteins (0.976–0.990), respectively.

Correlation among latent variables

Pearson correlation coefficients were calculated to understand the relationships among latent variables (Figure 6). Negative correlations were observed between mirna1 and mirna2 (−0.47), REA and prot4 (−0.33), REA and prot2 (−0.3), carcass and prot4

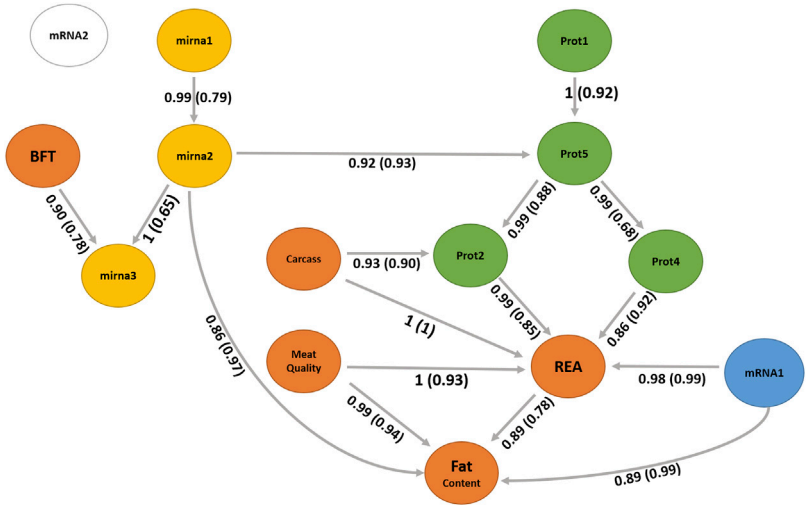


FIGURE 7
Bayesian network between latent variables based on the score-based (hill climbing and tabu) algorithms. The quality of the structure was evaluated by bootstrap resampling and model averaging across 1,000 replications. Orange nodes: phenotype latent variables; yellow nodes: miRNA latent variables; green nodes: protein latent variables; blue node: gene expression of mRNA1 module (WGCNA); and white node: gene expression of mRNA2 module (WGCNA). The labels of the arcs correspond to the strength and direction (in parenthesis).

(−0.31), carcass and prot2 (−0.28), and BFT and mirna3 (−0.25). Positive correlations are observed between all protein factors; meat quality and fat content (0.71), fat content and carcass (0.74), fat content and REA (0.76), and mirna2 and mirna3 (0.59). The latent variables REA and carcass correlated at 0.996. These results suggest that protein levels might have a negative impact on carcass, REA, and fat content factors.

Bayesian network

A BN was used to infer the interrelationships between latent variables. The BN algorithm learned with the most favorable network score in terms of BIC (1801.31) and BGe (1903.46) was the score-based hill climbing algorithm (Figure 7). The structure of BN was refined by model averaging with 1,000 networks from bootstrap resampling to reduce the impact of local optimal structures. The labels of the arcs measure the percentage of the uncertainty, corresponding to strength and direction (in parenthesis). The strength measures the frequency of the arc presented among all 1,000 networks from the bootstrapping replicates and the direction is the frequency of the direction shown conditionally in the presence of the arc.

We observed no difference in the structures between the two score-based algorithms used, the hill climbing and tabu. The two score-based algorithms produced a greater number of edges than the hybrid algorithms. The hill climbing algorithm produced 17 directed connections from the 14 latent variables.

Discussion

We integrated a multi-omic dataset with production, carcass, and meat quality traits and explored non-conventional relationships that led to new hypotheses in the meat quality field. Here, we applied EFA, BCFA, and BN to infer interrelationships among latent variables underlying complex traits and omic data. First, EFA and BCFA were used to reduce the dimensions of datasets by constructing latent variables and estimating their factor scores (de los Campos and Gianola, 2007). These latent variables represent more straightforward biological meanings than the original features measured in a population (Yu et al., 2019). Then, we applied a BN to understand the interrelationships among the latent variables (Neapolitan and others, 2004). We generated a network with 14 latent variables involving 17 directed connections. Moreover, this approach elucidated both direct and indirect relationships among latent variables. However, a precaution is essential to interpret the network as a causal relationship because causal statements require more assumptions (Pearl, 2009).

Yu et al. (2019) and Momen et al. (2021) applied a similar approach to obtain genetic insights on rice and wheat complex traits. Leal-Gutiérrez et al. (2018) studied the potential of using latent variables, obtained by structural equation analysis, on carcass and meat quality traits in beef cattle. They reduced the complexity of the data and reported biological mechanisms, such as postmortem proteolysis of structural proteins and cellular compartmentalization, cellular proliferation and

differentiation of adipocytes, and fat deposition. In recent work, Yu, et al. (2020) applied factor analysis to beef cattle behavior to better understand latent factors underlying temperament traits.

Biological meaning of latent variables and their relationships

The latent variable for the carcass, mainly composed of carcass cold weight, carcass depth, and fat kidney content (Supplementary Table S2), can be interpreted as the overall representation of the carcass, a higher value indicates a larger and heavier carcass. Its direct and indirect relationships with the latent variable REA (Figure 7) suggest the positive impact of the carcass yield on the ribeye area. Aass (1996) reported a positive phenotypic correlation (0.26) between the carcass depth and ribeye area corroborating our findings. Dinkel and Busch (1973) also estimated a positive genetic correlation (0.32) between growth rate and carcass yield, impacting the ribeye area positively.

The carcass has a relationship with the latent variable prot2 that also impacts REA. The latent variable prot2 is composed of 10 proteins (Supplementary Tables S2, S4), including UQCRC2 related with proteolysis (GO:0006508); ATP5F1A and ATP5F1B related with ATP synthesis coupled proton transport (GO:0042776 and GO:0015986); TNNT1 and TRIM72 related with muscle contraction (GO:0006936), regulation of muscle contraction (GO:0006937), sarcomere organization (GO:0045214), and muscle organ development (GO:0007517); GOT1 and GOT2 related with aspartate biosynthetic and catabolic processes (GO:0006532, GO:0006533), cellular response to insulin stimulus (GO:0032869), fatty acid homeostasis (GO:0055089), glutamate catabolic process to aspartate (GO:0019550), glycerol biosynthetic process (GO:0006114), and oxaloacetate metabolic process (GO:0006107); and MDH1 and MDH2 related with the carbohydrate metabolic process (GO:0005975), malate metabolic process (GO:0006108), NADH metabolic process (GO:0006734), oxaloacetate metabolic process (GO:0006107), tricarboxylic acid cycle (GO:0006099), and aerobic respiration (GO:0009060).

The ATP synthase F (0) complex subunit B1 (ATP5F1) has been positively correlated with meat color parameter a*, which impacts meat discoloration (Yu et al., 2017). Our findings show an indirect relationship between prot2 and the fat content latent variable that includes the parameter a*. Although prot2 and meat quality are not directly connected (Figure 7), both impact REA. However, prot2 is mainly composed of enzymes involved with energy metabolism that have been reported as putative candidate proteins for meat tenderness. The aspartate aminotransferase (GOT1) has been considered a putative candidate protein usable for meat tenderness prediction (Boudon et al., 2020). Rodrigues et al. (2017) reported that Nellore cattle have a higher abundance of malate dehydrogenase (MDH1) compared to Angus. This enzyme is important in gluconeogenesis, catalyzes the oxidation of malate to oxaloacetate, and is a relevant player in meat quality characteristics because this enzyme is involved in energy metabolism and affects how pH drops, changing the conversion of muscle to meat (Rodrigues et al., 2017). The ubiquinol-cytochrome C reductase core protein 2 (UQCRC2) gene, which is an important energy promoter for the development of cell functions was reported as up-regulated in a study analyzing gene expression on tough beef groups compared to the tender group in Nellore cattle (Muniz et al., 2021). The degradation of troponin T1 (TNNT1) proteins during post-mortem has been associated with meat tenderness (Zakrys-Waliwander et al., 2012; Contreras-Castillo et al., 2016; Wright et al., 2018).

The prot4 latent variable also shows a relationship with REA, which has two proteins (Supplementary Tables S2, S4) and includes the TNNI2 and MYH4. Troponin I, fast-twitch isoform (TNNI2) is a subunit of the troponin complex and plays a role in calcium regulation during muscle contraction and relaxation. The *TNNI2* gene was associated with pH, meat color value, and intramuscular fat content in pigs (Yang et al., 2010). The myosin heavy chains are relevant to muscle contraction velocity and power, MYH4 is one of the isoforms associated with IIb fibers types (Cho et al., 2016) and myotube hypertrophy in beef cattle (Bordbar et al., 2020). Our findings suggest new hypotheses of the impact of these proteins of prot2 could affect the REA and carcass traits.

The latent variable meat quality composed of shear force, myofibrillar fragmentation index, and the color parameters L* and b* can be interpreted as the overall representation of meat tenderness, and lower levels of this factor indicate more tender meat. It has a direct relationship with the latent variable REA and fat content. The relationship among tenderness, REA, and fat content has been discussed in the literature (Dinkel and Busch, 1973; Bonin et al., 2020). The mRNA1 latent variable has a relationship with REA and fat content. The mRNA1 factor is composed of the genes *LTN1, NFIA, ATP11B, FILIP1, RANBP2, N4BP2*, and *CERT1*. The nuclear factor IA gene (*NFIA*) has been studied indicating the potential to stimulate lipid accumulation in cattle (Chen H.-J. et al., 2018). According to the enrichment analysis (Supplementary Tables S5), these genes have been associated with an important cholesterol pathway called cholesterol and sphingolipid transport. Examples are the *RANBP2* gene which is associated with proteolysis and the *CERT1* gene which is related to intracellular cholesterol transport and sphingolipid metabolism. A further investigation is necessary to understand these relationships with REA or fat content.

The latent variable mirna3 is a child node of BFT and mirna2. The miRNAs are small RNA molecules that inhibit

translation or induce degradation of protein-coding mRNAs that contain complementary sequences to miRNAs. mirna3 is constituted by three miRNAs, namely, bta. let.7g, bta. miR.26b, and bta. miR.423.5p. bta. let.7 g was found in studies related with lactation and infection in cattle (Ma et al., 2019; Rani et al., 2020). mirna2 is composed of two miRNAs, namely, bta. let.7e and bta. miR.339b. Gu et al. (2007) identified the expression of bta. let.7e on adipose tissue in cattle. bta. miR.339b was found in studies related to fatty acid metabolism and lactation (Do et al., 2017; Palombo et al., 2018; Poleti et al., 2018). mirna2 has a direct relationship with fat content (Figure 7). Further studies are necessary to better understand the functions of mirna2 and its association with fat metabolism in beef cattle.

The latent variable prot5 is composed of two isoforms of nebulin (NEBU) which are important structural components involved in meat aging (Koohmaraie et al., 1984; Ouali et al., 1995). Post-mortem degradation of nebulin has been associated with meat tenderness in cattle in which animals with a lower degradation have less tender meat (Anderson and Parrish, 1989; Wu et al., 2014). The prot5 latent variable is an important node that has relationships with prot2 and prot4 and has an indirect relationship with REA and fat content.

The generated network identified interomic relationships, bringing simplicity without losing complexity. This is one of the challenges found in studies of this nature. Often a methodology used ends up providing the interpretation of unfeasible results, which was not in our approach. Additional investigations are essential to understand the relationships of molecules and phenotypes on latent variables REA, prot2, prot4, prot5, mRNA1, carcass, mirna3, mirna2, and fat content. The network demonstrated a relationship between miRNAs and nebulin protein isoforms that will not be found in studies using single or multi-level omics. Finally, REA appears as a central node in the network, influenced by carcass, prot2, prot4, and meat quality, suggesting that REA is a good indicator phenotype for meat quality because it can be easily measured during slaughter or by ultrasonography.

Conclusion

The FA identified latent variables, decreasing the dimensionality and complexity of data. The BN analysis was capable of identifying interrelationships among latent variables from different types of data, allowing the integration of different types of omic data and complex traits. The EFA, BCFA, and BN approaches can be used to generate new hypotheses on molecular research in the meat quality area, by integrating different types of data and exploring non-conventional relations.

Data availability statement

The mRNA datasets supporting the conclusion of this article are available in the European Nucleotide Archive (ENA) repository (EMBL-EBI), under accession nos. PRJEB13188, PRJEB10898, PRJEB15314, and PRJEB19421 (https://www.ebi.ac.uk/ena/browser/view/). The miRNA dataset of this article is available in the European Nucleotide Archive (ENA) repository (EMBL-EBI), under accession no. PRJEB42280. The protein data presented in the study are publicly available. These data can be found at: https://doi.org/10.1016/j.dib.2018.06.004. Any other relevant data are available from the authors upon reasonable request.

Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee Guidelines from EMBRAPA (CEUA 01/2013).

Author contributions

FN performed the bioinformatics and data analysis and drafted the manuscript. LC and GM conceived and conducted this study and supervised FN in analytical and data analysis and revised the manuscript. LR and LC designed the experimental study. AC and MP conducted the sample collection and the omic extraction protocols. MM and HY contributed to the data analysis. BP contributed to the functional enrichment analysis. All the authors have read and approved the final manuscript.

Funding

The experiments were financially supported by "Fundação de Amparo à Pesquisa do Estado de São Paulo," Brazil (FAPESP grants 12/23638-8 and 19/04089-2). This work was conducted while FJN was visiting Virginia Polytechnic Institute and State University, supported by the Coordinating Agency for Advanced Training of Graduate Personnel (CAPES grant: 88887.367967/2019-00).

Acknowledgments

LC, GM, and LR are recipients of CNPq productivity scholarships (304353/2019-1, 310714/2020-6, and 303754/2016-8, respectively).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.948240/full#supplementary-material

References

Aass, L. (1996). Variation in carcass and meat quality traits and their relations to growth in dual purpose cattle. *Livest. Prod. Sci.* 46, 1–12. doi:10.1016/0301-6226(96)00005-X

Adzitey, F. (2011). Effect of pre-slaughter animal handling on carcass and meat quality. *Int. Food Res. J.* 18.

Anderson, T. J., and Parrish, F. C. (1989). Postmortem degradation of titin and nebulin of beef steaks varying in tenderness. *J. Food Sci.* 54, 748–749. doi:10.1111/j.1365-2621.1989.tb04695.x

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*.

Bonin, M. de N., da Luz e Silva, S., Bünger, L., Ross, D., Feijó, G. L. D., da Costa Gomes, R., et al. (2020). Predicting the shear value and intramuscular fat in meat from Nellore cattle using Vis-NIR spectroscopy. *Meat Sci.* 163, 108077. doi:10.1016/j.meatsci.2020.108077

Bordbar, F., Jensen, J., Du, M., Abied, A., Guo, W., Xu, L., et al. (2020). Identification and validation of a novel candidate gene regulating net meat weight in Simmental beef cattle based on imputed next-generation sequencing. *Cell Prolif.* 53, e12870. doi:10.1111/cpr.12870

Boudon, S., Ounaissi, D., Viala, D., Monteils, V., Picard, B., and Cassar-Malek, I. (2020). Label free shotgun proteomics for the identification of protein biomarkers for beef tenderness in muscle and plasma of heifers. *J. Proteomics* 217, 103685. doi:10.1016/j.jprot.2020.103685

Carvalho, M. E., Gasparin, G., Poleti, M. D., Rosa, A. F., Balieiro, J. C. C., Labate, C. A., et al. (2014). Heat shock and structural proteins associated with meat tenderness in Nellore beef cattle, a *Bos indicus* breed. *Meat Sci.* 96, 1318–1324. doi:10.1016/j.meatsci.2013.11.014

Cerny, B. A., and Kaiser, H. F. (1977). A study of A measure of sampling adequacy for factor-analytic correlation matrices. *Multivar. Behav. Res.* 12, 43–47. doi:10.1207/s15327906mbr1201_3

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, 961. doi:10.1186/s12864-016-3306-x

Cesar, A. S., Regitano, L. C., Mourão, G. B., Tullio, R. R., Lanna, D. P., 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

Chen, H.-J., Ihara, T., Yoshioka, H., Itoyama, E., Kitamura, S., Nagase, H., et al. (2018a). Expression levels of brown/beige adipocyte-related genes in fat depots of vitamin A-restricted fattening cattle1. *J. Anim. Sci.* 96, 3884–3896. doi:10.1093/jas/sky240

Chen, Y., Mccarthy, D., Ritchie, M., Robinson, M., and Smyth, G. (2018b). *edgeR : differential expression analysis of digital gene expression data User ' s Guide*, 1–110.

Cho, E., Lee, K., Kim, J., Lee, S., Jeon, H., Lee, S., et al. (2016). Association of a single nucleotide polymorphism in the 5' upstream region of the porcine myosin heavy chain 4 gene with meat quality traits in pigs. *Anim. Sci. J.* 87, 330–335. doi:10.1111/asj.12442

Choi, T. (2015). Bayesian networks with examples in R. *Biometrics* 71, 864–865. doi:10.1111/biom.12369

Contreras-Castillo, C. J., Lomiwes, D., Wu, G., Frost, D., and Farouk, M. M. (2016). The effect of electrical stimulation on post mortem myofibrillar protein degradation and small heat shock protein kinetics in bull beef. *Meat Sci.* 113, 65–72. doi:10.1016/j.meatsci.2015.11.012

de Lima, A. O., Koltes, J. E., Diniz, W. J. S., de Oliveira, P. S. N., Cesar, A. S. M., Tizioto, P. C., et al. (2020). Potential biomarkers for feed efficiency-related traits in nelore cattle identified by Co-expression network and integrative genomics analyses. *Front. Genet.* 11, 189. doi:10.3389/fgene.2020.00189

de los Campos, G., and Gianola, D. (2007). Factor analysis models for structuring covariance matrices of additive genetic effects: A bayesian implementation. *Genet. Sel. Evol.* 39, 481–494. doi:10.1186/1297-9686-39-5-481

Dinkel, C. A., and Busch, D. A. (1973). Genetic parameters among production, carcass composition and carcass quality traits of beef cattle. *J. Anim. Sci.* 36, 832–846. doi:10.2527/jas1973.365832x

Do, D. N., Li, R., Dudemaine, P.-L., and Ibeagha-Awemu, E. M. (2017). MicroRNA roles in signalling during lactation: An insight from differential expression, time course and pathway analyses of deep sequence data. *Sci. Rep.* 7, 44605. doi:10.1038/srep44605

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinforma.* 51, 1–11. doi:10.1002/0471250953.bi1114s51

Filho, L. (2000). *Pecuária da carne bovina*. first edit. São Paulo: Embrapa.

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

Geromanos, S. J., Vissers, J. P. C., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V., et al. (2009). The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* 9, 1683–1695. doi:10.1002/pmic.200800562

Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* 28, 1694–1703. doi:10.1002/jssc.200500116

Gu, Z., Eleswarapu, S., and Jiang, H. (2007). Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland. *FEBS Lett.* 581, 981–988. doi:10.1016/j.febslet.2007.01.081

Guerrero, A., Valero, M. V., Campo, M. M., and Sañudo, C. (2013). Some factors that affect ruminant meat quality: From the farm to the fork. Review. *Acta Sci. Anim. Sci.* 35. doi:10.4025/actascianimsci.v35i4.21756

Hopkins, D. ., Littlefield, P., and Thompson, J. (2000). A research note on factors affecting the determination of myofibrillar fragmentation. *Meat Sci.* 56, 19–22. doi:10.1016/S0309-1740(00)00012-7

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi:10.1007/BF02289447

Horwitz, W. (2000). *Official methods of analysis of AOAC International*. 17th ed. Gaithersburg, Md.

Huang, S., Chaudhary, K., and Garmire, L. X. (2017). More is better: Recent progress in multi-omics data integration methods. *Front. Genet.* 8, 84–12. doi:10.3389/fgene.2017.00084

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.* 2, 343–372. doi:10.1146/annurev.genom.2.1.343

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200. doi:10.1007/BF02289233

Kappeler, B. I. G., Regitano, L. C. A., Poleti, M. D., Cesar, A. S. M., Moreira, G. C. M., Gasparin, G., et al. (2019). MiRNAs differentially expressed in skeletal muscle of animals with divergent estimated breeding values for beef tenderness. *BMC Mol. Biol.* 20, 1. doi:10.1186/s12867-018-0118-3

Koohmaraie, M., Kennick, W. H., Elgasim, E. A., and Anglemier, A. F. (1984). Effects of postmortem storage on muscle protein degradation: Analysis by SDS-polyacrylamide gel electrophoresis. *J. Food Sci.* 49, 292–293. doi:10.1111/j.1365-2621.1984.tb13732.x

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28. doi:10.18637/jss.v028.i05

Lalli, P. M., Corilo, Y. E., Fasciotti, M., Riccio, M. F., de Sa, G. F., Daroda, R. J., et al. (2013). Baseline resolution of isomers by traveling wave ion mobility mass spectrometry: Investigating the effects of polarizable drift gases and ionic charge distribution. *J. Mass Spectrom.* 48, 989–997. doi:10.1002/jms.3245

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

Leal-Gutiérrez, J. D., Rezende, F. M., Elzo, M. A., Johnson, D., Peñagaricano, F., and Mateescu, R. G. (2018). Structural equation modeling and whole-genome scans uncover chromosome regions and enriched pathways for carcass and meat quality in beef. *Front. Genet.* 9, 532. doi:10.3389/fgene.2018.00532

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

Ma, S., Tong, C., Ibeagha-Awemu, E. M., and Zhao, X. (2019). Identification and characterization of differentially expressed exosomal microRNAs in bovine milk infected with *Staphylococcus aureus*. *BMC Genomics* 20, 934. doi:10.1186/s12864-019-6338-1

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10. doi:10.14806/ej.17.1.200

Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., and Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.* 17, 628–641. doi:10.1093/bib/bbv108

Merkle, E. C., and Rosseel, Y. (2018). Blavaan : Bayesian structural equation models via parameter expansion. *J. Stat. Softw.* 85. doi:10.18637/jss.v085.i04

MetaCore (2021). MetaCore login | clarivate. Available at: https://portal.genego.com/(Accessed June 8, 2021).

Misra, B. B., Langefeld, C., Olivier, M., and Cox, L. A. (2018). Integrated omics: Tools, advances and future approaches. *J. Mol. Endocrinol.* 62, R21–R45. doi:10.1530/jme-18-0055

Momen, M., Bhatta, M., Hussain, W., Yu, H., and Morota, G. (2021). Modeling multiple phenotypes in wheat using data-driven genomic exploratory factor analysis and Bayesian network learning. *Plant Direct* 5, e00304. doi:10.1002/pld3.304

Muniz, M. M. M., Fonseca, L. F. S., dos Santos Silva, D. B., de Oliveira, H. R., Baldi, F., Chardulo, A. L., et al. (2021). Identification of novel mRNA isoforms associated with meat tenderness using RNA sequencing data in beef cattle. *Meat Sci.* 173, 108378. doi:10.1016/j.meatsci.2020.108378

Nascimento, M. L., Souza, A. R. D. L., Chaves, A. S., Cesar, A. S. M., Tullio, R. R., Medeiros, S. R., et al. (2016). Feed efficiency indexes and their relationships with carcass, non-carcass and meat quality traits in Nellore steers. *Meat Sci.* 116, 78–85. doi:10.1016/j.meatsci.2016.01.012

Neapolitan, R. E. (2004). *Learning bayesian networks*. Upper Saddle River, NJ: Pearson Prentice Hall.

Njisane, Y. Z., and Muchenje, V. (2016). Farm to abattoir conditions, animal factors and their subsequent effects on cattle behavioural responses and beef quality — a review. *Asian-Australas. J. Anim. Sci.* 30, 755–764. doi:10.5713/ajas.16.0037

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

Ouali, A., Demeyer, D., and Smulders, F. (1995). "Editorial," in *Tissue proteinases and regulation of protein degradation as related to meat quality* (Nijmegen: ECCEAMST), V–VI.

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., et al. (2018). Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. *J. Dairy Sci.* 101, 11004–11019. doi:10.3168/jds.2018-14413

Pearce, K. L., Rosenvold, K., Andersen, H. J., and Hopkins, D. L. (2011). Water distribution and mobility in meat during the conversion of muscle to meat and ageing and the impacts on fresh meat quality attributes — a review. *Meat Sci.* 89, 111–124. doi:10.1016/j.meatsci.2011.04.007

Pearl, J. (2009). *Causality*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803161

Poleti, M. D., Regitano, L. C. A., Souza, G. H. M. F., Cesar, A. S. M., Simas, R. C., Silva-Vignato, B., et al. (2018). Longissimus dorsi muscle label-free quantitative proteomic reveals biological mechanisms associated with intramuscular fat deposition. *J. Proteomics* 179, 30–41. doi:10.1016/j.jprot.2018.02.028

Rani, P., Onteru, S. K., and Singh, D. (2020). Genome-wide profiling and analysis of microRNA expression in buffalo milk exosomes. *Food Biosci.* 38, 100769. doi:10.1016/j.fbio.2020.100769

Revelle, W. R. (2017). *psych: Procedures for personality and psychological research*.

Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015a). Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* 16, 85–97. doi:10.1038/nrg3868

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

Rodin, A. S., and Boerwinkle, E. (2005). Mining genetic epidemiology data with Bayesian networks I: Bayesian networks and example application (plasma apoE levels). *Bioinformatics* 21, 3273–3278. doi:10.1093/bioinformatics/bti505

Rodrigues, R. T. de S., Chizzotti, M. L., Vital, C. E., Baracat-Pereira, M. C., Barros, E., Busato, K. C., et al. (2017). Differences in beef quality between angus (*Bos taurus* taurus) and Nellore (*Bos taurus* indicus) cattle through a proteomic and phosphoproteomic approach. *PLoS One* 12, e0170294. doi:10.1371/journal.pone.0170294

Rust, S. R., Price, D. M., Subbiah, J., Kranzler, G., Hilton, G. G., Vanoverbeke, D. L., et al. (2008). Predicting beef tenderness using near-infrared spectroscopy. *J. Anim. Sci.* 86, 211–219. doi:10.2527/jas.2007-0084

Scutari, M. (2010). Learning bayesian networks with the**bnlearn***R*Package. *J. Stat. Softw.* 35, 1–22. doi:10.18637/jss.v035.i03

Silva, W. M., Carvalho, R. D., Soares, S. C., Bastos, I. F., Folador, E. L., Souza, G. H., et al. (2014). Label-free proteomic analysis to confirm the predicted proteome of Corynebacterium pseudotuberculosis under nitrosative stress mediated by nitric oxide. *BMC Genomics* 15, 1065. doi:10.1186/1471-2164-15-1065

Souza, G. H. M. F., Guest, P. C., and Martins-de-Souza, D. (2017). LC-MSE, multiplex MS/MS, ion mobility, and label-free quantitation in clinical proteomics. *Methods Mol. Biol.* 1546, 57–73. doi:10.1007/978-1-4939-6730-8_4

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

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

Wheeler, T. L., Cundiff, L. V., Shackelford, S. D., and Koohmaraie, M. (2005). Characterization of biological types of cattle (Cycle VII): Carcass, yield, and longissimus palatability traits. *J. Anim. Sci.* 83, 196–207. doi:10.2527/2005.831196x

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

Wright, S. A., Ramos, P., Johnson, D. D., Scheffler, J. M., Elzo, M. A., Mateescu, R. G., et al. (2018). Brahman genetics influence muscle fiber properties, protein degradation, and tenderness in an Angus-Brahman multibreed herd. *Meat Sci.* 135, 84–93. doi:10.1016/j.meatsci.2017.09.006

Wu, G., Farouk, M. M., Clerens, S., and Rosenvold, K. (2014). Effect of beef ultimate pH and large structural protein changes with aging on meat tenderness. *Meat Sci.* 98, 637–645. doi:10.1016/j.meatsci.2014.06.010

Yang, H., Xu, Z. Y., Lei, M. G., Li, F. E., Deng, C. Y., Xiong, Y. Z., et al. (2010). Association of 3 polymorphisms in porcine troponin I genes (TNNI1 andTNNI2) with meat quality traits. *J. Appl. Genet.* 51, 51–57. doi:10.1007/BF03195710

Yu, H., Campbell, M. T., Zhang, Q., Walia, H., and Morota, G. (2019). Genomic Bayesian confirmatory factor analysis and Bayesian network to characterize a wide spectrum of rice phenotypes. *G3* 9, 1975–1986. doi:10.1534/g3.119.400154

Yu, H., Morota, G., Celestino, E. F., Dahlen, C. R., Wagner, S. A., Riley, D. G., et al. (2020). Deciphering cattle temperament measures derived from a four-platform standing scale using genetic factor Analytic modeling. *Front. Genet.* 11, 599. doi:10.3389/fgene.2020.00599

Yu, Q., Wu, W., Tian, X., Hou, M., Dai, R., and Li, X. (2017). Unraveling proteome changes of Holstein beef M. semitendinosus and its relationship to meat discoloration during post-mortem storage analyzed by label-free mass spectrometry. *J. Proteomics* 154, 85–93. doi:10.1016/j.jprot.2016.12.012

Yu, Y.-Q., Gilar, M., Lee, P. J., Bouvier, E. S. P., and Gebler, J. C. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* 75, 6023–6028. doi:10.1021/ac0346196

Zakrys-Waliwander, P. I., O'Sullivan, M. G., O'Neill, E. E., and Kerry, J. P. (2012). The effects of high oxygen modified atmosphere packaging on protein oxidation of bovine M. longissimus dorsi muscle during chilled storage. *Food Chem. x.* 131, 527–532. doi:10.1016/j.foodchem.2011.09.017

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fvets.2022.933850

OPEN ACCESS

EDITED BY
Ana Fabrícia Braga Magalhães,
Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

REVIEWED BY
Hui Li,
Guangxi University, China
Sayed Haidar Abbas Raza,
Northwest A&F University, China
Zhuanjian Li,
Henan Agricultural University, China
Yinghui Ling,
Anhui Agricultural University, China

*CORRESPONDENCE
Tieshan Xu
xutieshan760412@163.com

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

RECEIVED 01 May 2022
ACCEPTED 28 September 2022
PUBLISHED 24 October 2022

CITATION
Gu L, Zhang S, Li B, Jiang Q, Xu T, Huang Y, Lin D, Xing M, Huang L, Zheng X, Wang F, Chao Z and Sun W (2022) m6A and miRNA jointly regulate the development of breast muscles in duck embryonic stages. *Front. Vet. Sci.* 9:933850. doi: 10.3389/fvets.2022.933850

m6A and miRNA jointly regulate the development of breast muscles in duck embryonic stages

Lihong Gu[1,2], Shunjin Zhang[3], Boling Li[4], Qicheng Jiang[5], Tieshan Xu[1]*, Yongzhen Huang[3], Dajie Lin[2], Manping Xing[2,6], Lili Huang[2,6], Xinli Zheng[2,6], Feng Wang[2], Zhe Chao[2,6] and Weiping Sun[1]

[1]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China, [2]Institute of Animal Science and Veterinary Medicine, Hainan Academy of Agricultural Sciences, Haikou, China, [3]College of Animal Science and Technology, Northwest A&F University, Xianyang, China, [4]The Hainan Animal Husbandry Technology Promotion Station, Haikou, China, [5]School of Life Science, Hainan University, Haikou, China, [6]Key Laboratory of Tropical Animal Breeding and Disease Research, Haikou, China

N6-methyladenosine (m6A) is an abundant internal mRNA modification and plays a crucial regulatory role in animal growth and development. In recent years, m6A modification has been found to play a key role in skeletal muscles. However, whether m6A modification contributes to embryonic breast muscle development of Pekin ducks has not been explored. To explore the role of m6A in embryonic breast muscle development of ducks, we performed m6A sequencing and miRNA sequencing for the breast muscle of duck embryos on the 19th (E19) and 27th (E27) days. A total of 12,717 m6A peaks were identified at E19, representing a total of 7,438 gene transcripts. A total of 14,703 m6A peaks were identified, which overlapped with the transcripts of 7,753 genes at E27. Comparing E19 and E27, we identified 2,347 differential m6A peaks, which overlapped with 1,605 m6A-modified genes (MMGs). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses revealed that MMGs were enriched in multiple muscle- or fat-related pathways, which was also revealed from our analysis of differentially expressed genes (DEGs). Conjoint analysis of m6A-seq and RNA-seq data showed that pathways related to β-oxidation of fatty acids and skeletal muscle development were significantly enriched, suggesting that m6A modification is involved in the regulation of fat deposition and skeletal muscle development. There were 90 upregulated and 102 downregulated miRNAs identified between the E19 and E27 stages. Through overlapping analysis of genes shared by MMGs and DEGs and the targets of differentially expressed miRNAs (DEMs), we identified six m6A-mRNA-regulated miRNAs. Finally, we found that m6A modification can regulate fat deposition and skeletal muscle development. In conclusion, our results suggest that m6A modification is a key regulator for embryonic breast muscle development and fat deposition of ducks by affecting expressions of mRNAs and miRNAs. This is the first study to comprehensively characterize the m6A patterns in the duck transcriptome.

These data provide a solid basis for future work aimed at determining the potential functional roles of m6A modification in adipose deposition and muscle growth.

KEYWORDS

ducks, embryo, breast muscles, m6A sequencing, miRNAs sequencing

Introduction

More than 150 chemical modifications to RNA have been described (1), and these structural modifications play regulatory roles by affecting gene expression. In recent years, with the identification of enzymes capable of reversing N6-methyladenosine (m6A) and the development of transcriptome-wide sequencing methods to map modified sites (2–5), the prevalence and functional significance of internal mRNA modifications have been recognized. Therefore, there is a renewed interest in the biological function of RNA modification.

N6-methyladenosine was first discovered in 1974 and refers to the RNA methylation modification on the sixth N atom of base A with the active adenosine acid as the methyl donor (6). m6A is the most abundant internal mRNA modification and has been observed in various species, accounting for over 80% of all RNA base methylations (7–9). The enzymes that modify m6A methylation include "writers," "erasers," and "readers," which refer to methylated transferases, demethylases, and methylated reading proteins, respectively (10). Methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14), regulated by the association of a subunit protein Wilms tumor 1-associated protein (WTAP), belong to methylated transferase and can form complexes to catalyze the deposition of m6A in mammalian mRNA (11, 12). AlkB homolog 5 (ALKBH5) and fat mass and obesity-associated (FTO) protein act as m6A demethylases to remove methyl from target regions (13, 14), while heterogeneous nuclear ribonucleoprotein and YTH domain-containing RNA-binding protein act as m6A-methylated reading proteins (15).

N6-methyladenosine is closely involved in the regulation of gene expression, RNA transcription, translation, shearing, degradation, and nuclear transportation (16, 17). More importantly, m6A methylation modification plays an essential role in animal growth and development. METTL3 knockdown inhibited myoblast proliferation and myogenic differentiation, whereas METTL3 overexpression promoted these processes (18). A complete transcriptome map of m6A was obtained by transcriptome sequencing of muscle tissue from three different pig breeds, and m6A was found to be widely distributed in muscle tissue (19). However, to our knowledge, no study has addressed m6A modification in the breast muscle tissues of ducks.

China is a major producer and consumer of ducks in the world, and in 2019, duck production and consumption in China accounted for about 75% of the world's duck stock, according to the Food and Agriculture Organization (FAO). Pekin duck is a famous meat breed for its fast growth rate. Therefore, a study on the regulation mechanism of breast muscle development is crucial for improving the meat yield of Pekin duck and is also the basis of breeding a new lean meat type of Pekin duck. Our previous research showed that the 19th day of hatching (E19) is the fastest point of breast muscle development—as well as the crucial transition point for breast muscle development during the embryonic stage of Pekin ducks—and that the weight of the breast muscles was largely constant from E19 to E27 (20). Subsequently, miRNA and mRNA patterns of breast muscles at the E13 (13th day of hatching), E19, and E27 (27th day of hatching) stages were studied (21, 22), and candidates that may play key roles in the breast muscle development of Pekin duck were identified. However, no studies have shown whether m6A and miRNA expressions are different in E19 and 27 and the potential role of m6A in the breast muscle development of Pekin duck.

The aim of this study was to explore whether m6A and miRNA cooperatively regulate breast muscle differentiation in Pekin duck embryos. We conducted m6A sequencing on the breast muscle tissues of ducks at E19 and E27 to explore whether m6A modification existed in these two periods. We also performed miRNA-seq to explore whether some genes were regulated by both m6A and miRNA in Pekin duck embryos. We determined the distribution of m6A and miRNA regulation during breast muscle differentiation in Pekin duck embryos. The results of this study will offer a basis for unraveling the role of m6A modification and miRNA in breast muscle differentiation.

Materials and methods

Ethics approval

This study was approved by the Institute of Animal Science & Veterinary Medicine, Hainan Academy of Agricultural Sciences (IASVM-HAAS, Haikou, China; ethical approval reference number: IASVMHAAS-AE-202012), and followed the Regulations for the Administration of Affairs Concerning Experimental Animals of China.

Searching of duck homologous sequences in m6A RNA modification

The duck reference genome sequences (BGI_duck_1.0) and complete genome annotation GFF3 file were downloaded from the NCBI database, which ensured that genes affected by m6A methylation were accurately located. Related enzyme sequences in m6A RNA modification were searched by blasting the duck reference genome, and the duck sequences were compared with those of human.

Sample collection

The experimental duck embryos were obtained from the Z-type Pekin Duck Breeding Farm of the Beijing Institute of Animal Science, Chinese Academy of Agricultural Sciences. The eggs were incubated at a temperature of 37 ± 0.5°C and a humidity of 86–87%. The breast muscle samples were obtained from E13, E19, and E27 stages. The duck embryo and breast muscle were taken out and spun off under aseptic conditions. Three eggs at each stage were used to collect breast muscles. The left and right breast muscles were placed in separate centrifuge tubes, placed in liquid nitrogen, taken back to the laboratory, and stored in a −80°C cryogenic refrigerator. The left and right breast muscles from each duck embryo were used for m6A-seq and miRNA-seq sequencing analyses, respectively.

RNA extraction

Total RNA was isolated and purified using Trizol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. The RNA concentration and purity of each sample were quantified using a NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA). The RNA integrity was assessed by a Bioanalyzer 2100 (Agilent, CA, USA) with RIN number >7.0 and confirmed by electrophoresis with denaturing agarose gel. Poly (A) RNA was purified from 50 μg total RNA using Dynabeads Oligo (dT)25-61005 (Thermo Fisher, CA, USA). Then, the poly (A) RNA was fragmented into small pieces using a Magnesium RNA Fragmentation Module (NEB, cat. e6150, USA) at 86°C for 7 min.

Expression pattern of m6A-associated methylase

Total RNA of each breast muscle sample from three embryonic stages was isolated using Trizol reagent (Invitrogen, USA) following the manufacturer's instructions. The SYBR PrimeScript RT-PCR Kit (TaKaRa, Japan) was used for reverse transcription polymerase chain reaction (RT-PCR). The relative expression levels of *METTL3*, *METTL14*, *WATP*, *KIAA1429*, *FTO*, *ALKBH5*, *YTHDF1*, and *YTHDF2* were examined by quantitative RT-PCR (qRT-PCR) using reference gene GAPDH. The primer sequences are listed in Table 1.

TABLE 1 Primers information of genes.

Primer labels	Sequence (5′ − 3′)	Annealing temperature °C
GAPDH-F	CACACGAAGACAGTGGATG	60
GAPDH-R	GAGGCTGGGGATAATGTTCTG	
METTL3-F	GCTCCACCAGCCATAAACC	56
METTL3-R	TGAACTGCGCCACCACAT	
METTL14-F	TGAACAGTAAGGATGACCA	60
METTL14-R	TTGGAGCAGAGGTATCATAA	
WATP-F	TCCAGGAGAATCAAGAGC	53
WATP-R	CATTGCTTGGTCCGTTAG	
KIAA1429-F	TTCTTCTTGCCAGCCTATG	55
KIAA1429-R	ATCCCAGTGTATCCGAGTA'	
FTO-F	ACCTGCTGAAGAAACTTATGAT	60
FTO-R	TTGGTGAAGTGGTATTGCTAAT	
ALKBH5-F	GGAGGGTTACACCTACGGC	57
ALKBH5-R	CCTGATGGGTTTGAACTGGA	
YTHDF1-F	GACTCAACCACAGTATCAGA	60
YTHDF1-R	GTTACCAGTTCCTCCACTT	
YTHDF2-F	CTCTCACGGCTTCCTAAT	60
YTHDF2-R	CGCTTCTGTTGGTCTTATC	

Quantitative RT-PCR was carried out with an iCycler IQ5 Multicolor Real-Time PCR Detection System (Bio-Rad, USA). The qRT-PCR contained 1 μL of cDNA, 12.5 μL of SYBR Premix Ex-Taq, 10.5 μL of ddH_2O, and 0.5 μL of 10 pmol/μL forward and reverse primer (Table 1). The thermal cycling parameters were one cycle at 95°C for 30 s and 40 cycles at 95°C for 10 s and 60°C for 40 s. An 80-cycle melting curve analysis was performed after each PCR run to confirm product specificity, with one cycle at 95°C for 1 min, one cycle at 55°C, and then increasing temperature of 0.5°C for every 10 s until 95°C while fluorescence was continuously monitored. qRT-PCR analysis of each sample was repeated three times.

m6A immunoprecipitation (IP), library construction, and sequencing

The cleaved RNA fragments were incubated at 4°C for 2 h with an m6A-specific antibody (Synaptic Systems, cat. 202003, Germany) in IP buffer (50 mM Tris–HCl, 750 mM NaCl, and 0.5% Igepal CA-630). Then, the IP RNA was reverse transcribed to create the cDNA by SuperScript™ II reverse

transcriptase (Invitrogen, cat. 1896649, USA), which was then used to synthesize U-labeled second-stranded DNAs with *E. coli* DNA polymerase I (NEB, cat. m0209, USA), RNase H (NEB, cat. m0297, USA), and dUTP solution (Thermo Fisher, cat. R0133, USA). An A-base was then added to the blunt ends of each strand for ligating to the indexed adapters. Each adapter contained a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. Single- or dual-index adapters were ligated to the fragments, and the sample size selection was performed with AMPureXP beads. After the heat-labile UDG enzyme (NEB, cat. m0280, USA) treatment of the U-labeled second-stranded DNAs, the ligated products were amplified by PCR under the following conditions: denaturation at 95°C for 3 min, eight cycles of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, extension at 72°C for 30 s, and final extension at 72°C for 5 min. The average insert size for the final cDNA library was 300 ± 50 bp. Finally, the 2 × 150bp paired-end sequencing (PE150) was performed on an Illumina NovaSeq™ 6000 (LC-Bio Technology Co., Ltd., Hangzhou, China) following the vendor's recommended protocol.

Data analysis of m6A-seq and RNA-seq

The fastp tool (https://github.com/OpenGene/fastp) was used to remove the reads that contained adaptor contamination, low-quality bases, and undetermined bases with default parameters. Then, sequence quality of IP and Input samples was also verified using fastp. We used HISAT2 (http://daehwankimlab.github.io/hisat2) (23) to map reads to the reference genome of *Anas platyrhynchos*. The mapped reads of IP and Input libraries were provided to an R package exomePeak (https://bioconductor.org/packages/exomePeak), which can identify m6A peaks with bed or bigwig format files that can be adapted for visualization on the IGV software (http://www.igv.org). Peaks were examined by the Poisson distribution matrix with default parameters ($P < 0.05$). MEME (http://meme-suite.org) and HOMER (http://homer.ucsd.edu/homer/motif) were used for *de novo* and known motif findings, followed by localization of the motif with respect to peak summit. Called peaks were annotated by intersection with gene architecture using the R package ChIPseeker (https://bioconductor.org/packages/ChIPseeker). Then, StringTie (https://ccb.jhu.edu/software/stringtie) was used to determine the expression level for all mRNAs from Input libraries by calculating FPKM [total exon fragments/mapped reads (millions) × exon length (kB)]. The differentially expressed genes (DEGs) were selected with log2 (fold change) >1 or log2 (fold change) <-1 and *p*-value < 0.05 using the R package edgeR (https://bioconductor.org/packages/edgeR) (24). Gene enrichment analysis was performed by Gene Ontology (GO) (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.kegg.jp/).

Conjoint analysis of m6A-seq and RNA-seq data

To comprehensively study the roles of methylated m6A level and gene expression abundance, we performed correlation analyses of m6A-seq and RNA-seq data. Through the analyses of m6A-seq and RNA-seq data, we obtained differentially methylated m6A peaks in abundance and DEGs. We divided the differentially methylated m6A peaks into upregulated m6A sites (higher methylated m6A sites at E27 than those at E19) and downregulated m6A sites (higher methylated m6A sites at E19 than those at E27). Similarly, DEGs were also divided into upregulated genes (higher expression levels at E27 than those at E19) and downregulated genes (higher expression levels at E19 than those at E27). We overlapped up- and down-methylated m6A sites with up- and downregulated genes and then obtained the upregulated genes with upregulated methylated m6A sites (up–up), downregulated genes with upregulated methylated m6A sites (down–up), upregulated genes with downregulated methylated m6A sites (up–down), and downregulated genes with hypo-methylated m6A sites (down–down). GO and KEGG functional analyses for genes shared by DEGs and DMGs were performed to investigate the functions of these genes.

miRNA library construction, sequencing, and data analysis

The miRNA libraries were constructed with a similar method to Gu et al. (22). In brief, the isolated total RNA of each individual from E19 and E27 was used for the generation of the small RNA libraries where the population of recovered small RNAs, ranging in size from 18 to 30 nucleotides, was purified using 15% polyacrylamide gel. Then, 5′ adaptors (Illumina, USA) were ligated to the purified small RNAs, followed by purification of ligation products on Novex 15% TBE–urea gel. The 5′ ligation products were then ligated to 3′ adaptors (Illumina), and products with 5′ and 3′ adaptors were purified using Novex 10% TBE–urea gel (Invitrogen). Subsequently, reverse transcription reactions were performed using the RT primer, and PCRs were performed using the forward and reverse Illumina primers. The PCR product was purified by phenol/chloroform extraction and ethanol precipitation, and miRNA libraries were obtained. After purification and quality detection, miRNA libraries were sequenced on an Illumina Genome Analyzer (LC-Bio Technology Co., Ltd., Hangzhou, China), and raw reads were produced.

The raw reads were subjected to an in-house program, ACGT101-miR (LC Sciences, Houston, TX, USA), to remove adapter dimers, junk, low complexity, common RNA families (rRNA, tRNA, snRNA, and snoRNA), and repeats. Subsequently, unique sequences with lengths of 18–26 nucleotides were

mapped to specific species precursors in miRBase 22.0 by BLAST search to identify known miRNAs and novel 3p- and 5p-derived miRNAs. Length variation at both the 3′ and 5′ ends and only one mismatch inside of the sequence were allowed in the alignment. The unique sequences mapping to specific species mature miRNAs in hairpin arms were identified as known miRNAs. The unique sequences mapping to the other arm of known specific species precursor hairpin opposite to the annotated mature miRNA-containing arm were considered as novel 5p- or 3p-derived miRNA candidates.

The unmapped sequences were BLASTed against the specific genomes, and the hairpin RNA structures containing sequences were predicated from the flank 80 nt sequences using RNAfold software (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi). The criteria for secondary structure prediction were as follows: (1) number of nucleotides in one bulge in stem (≤12); (2) number of base pairs in the stem region of the predicted hairpin (≥16); (3) cutoff of free energy (kCal/mol ≤-15); (4) length of the hairpin (up and down stems + terminal loop ≥50); (5) length of the hairpin loop (≤20); (6) number of nucleotides in one bulge in the mature region (≤8); (7) number of biased errors in one bulge in the mature region (≤4); (8) number of biased bulges in the mature region (≤2); (9) number of errors in the mature region (≤7); (10) number of base pairs in the mature region of the predicted hairpin (≥12); and (11) percent of mature in stem (≥80).

Differentially expressed miRNAs were selected by the t-test method (http://en.wikipedia.org/wiki/Students_t-test), which compares the significance level of difference between the E19 and E27 stages. A heatmap was used to analyze the cluster pattern in different control sets with log10 values. Target genes of DEMs with a significant difference were predicted by the TargetScan algorithm (25–27) with default parameter and miRanda algorithm (28, 29) (Max_Energy<-10) according to the score standard. Finally, the overlapped genes predicted by both algorithms were deemed as the target genes of DEMs.

Integrative analysis of miRNA-seq, mRNA-seq, and m6A-seq data

Differentially expressed miRNAs were identified through the analysis of miRNA-seq data, and we predicted the targets of DEMs. Then, we overlapped the targets of DEMs and the genes shared by MMGs and DEGs (called m6A-mRNA-miRNA genes). Subsequently, we identified the miRNAs that were targeted by m6A-mRNA-miRNA genes (called m6A-mRNA-regulated miRNAs). Finally, we performed GO and KEGG functional enrichment analyses of targets of m6A-mRNA-regulated miRNAs to study their potential roles.

Availability of data

Sequences are available from GenBank with the Bioproject accession numbers SRR13051312–SRR13051329.

Results

Comparison of homologous genes of methylase

The sequences of m6A-methylated enzymes in ducks were obtained by blasting on NCBI. Through comparison with human homologous genes of methylases, we found that all duck *METTL14*, *FTO*, and *YTHDF1* genes had homologous genes in humans. All the E-values of these three genes were 0.0, and their max scores were 674, 605, and 802, respectively.

m6A modification levels and the expression of m6A RNA modification enzymes

We previously found that duck breast muscles grew much faster at E19 than those at E13 and E27. To explore whether m6A modification played a role in duck embryonic breast muscle development, we examined the levels of m6A in the total RNAs of breast muscles at E13, E19, and E27. Using the colorimetric m6A quantification strategy, we found that the m6A level at E19 was much higher than those at E13 and E27 (Figure 1A), which indicated that the expression of the total RNA methylation level of m6A in duck breast muscle was significantly higher in the period of vigorous proliferation and differentiation than that during the period of cessation of proliferation.

According to our previous transcriptome sequencing (RNA-seq) results, the expression of *METTL14*, *FTO*, and *YTHDF1* was significantly different among the E13, E19, and E27 stages ($P < 0.01$) (Figure 1B). Then, qRT-PCR was used to test the mRNA expression levels of several m6A RNA modification enzymes during different stages. The results revealed that the expression levels of *METTL14*, *FTO*, *YTHDF1*, and *YTHDF2* were higher in the breast at E19 than those at E27, while *WATP*, *KIAA1429*, and *ALKBH* were more highly expressed at E27 than at E19 (Figure 1C). Therefore, we speculated that the expression of methylase was significantly correlated with the proliferation and differentiation of duck breast muscle cells.

Transcriptome-wide m6A modification patterns

Through sequencing of m6A-seq and RNA-seq, 68.19–81.52 million clean reads were generated from each m6A-seq dataset

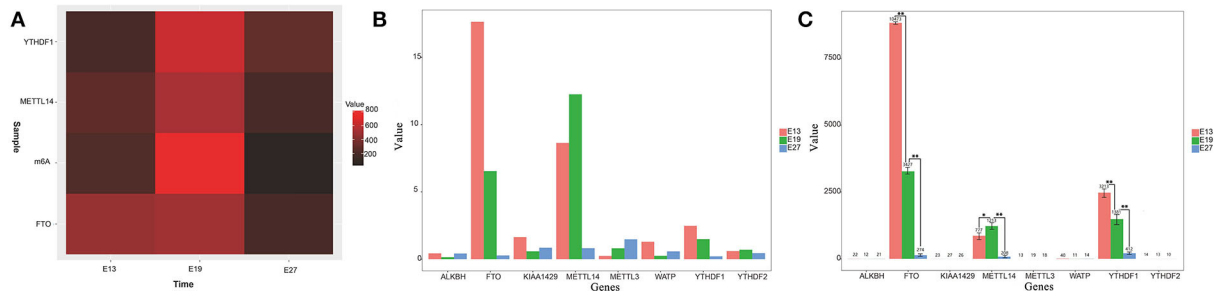


FIGURE 1
Gene expression levels of m6A methylation enzymes in different stages. **(A)** Heatmap of m6A modification levels of E13, E19, and E27 in duck embryonic breast muscles. **(B)** Gene expression levels of m6A methylation enzymes at E13, E19, and E27 from RNA-seq data. **(C)** Detection of gene expression levels of m6A methylation enzymes at E13, E19, and E27 detected by qRT-PCR. Statistically significant differences are indicated by $*p \leq 0.05$, $**p \leq 0.01$.

and 65.48–88.88 million clean reads were generated from the RNA-seq dataset (Table 2). HISAT2 was used to map reads to the genome of *Anas platyrhynchos* (BGI_duck_1.0) with default parameters. The percentages of mapped reads for m6A-seq ranged from 75.25 to 77.01% and for RNA-seq ranged from 74.45 to 77.62% (Supplementary Table S1).

To study the distribution of m6A peaks, we first detected the enrichment of reads near the transcriptome initiation site (transcription start sites, TSS). The results showed that the m6A peaks were enriched in the vicinity of transcription start sites (TSS) (Figure 2A). We then divided the duck genes into 5'UTR, 3'UTR, first exon, and other exons and found that the reads from m6A-IP samples are highly accumulated around the other exons (except the first exon) at both E19 (53.81%) and E27 (36.17%) stages (Figures 2B,C). For E19, the ratios of peaks were similar on the first exon (18.81%) and 3′UTR (18.69%), while peaks distributed at 5′UTR were the least (8.69%). For E27, the percentages of peaks on the first exon and the other exons were also similar, all accounting for about 36%. A total of 12,717 m6A peaks were identified by exomePeak at E19, representing the transcripts of 7,438 genes. At E27, 14,703 m6A peaks were identified, which correspond to the transcripts of 7,753 genes. There were 5,091 and 5,406 unique peaks for E19 and E27, but only 2,347 peaks were shared by E19 and E27, indicating their significant difference in global m6A modification patterns (Figure 2D).

Enrichment analysis of m6A-modificated genes

The abundance of the m6A peaks between the E19 and E27 stages was compared to identify the abundance differential peaks (Supplementary material 1). We found 2,347 differential m6A peaks between the E19 and E27 stages, which overlapped 1,605 genes (called m6A-modified genes, MMGs). The MMGs were mainly located at other exons and 3′UTR regions (Figure 3A). GO and KEGG analyses were then performed to explore the function of m6A-modified genes. GO analysis displayed that MMGs were mostly enriched in cellular component and molecular functions (Figure 3B). Moreover, there were 179 genes located in the nucleus. Furthermore, 169 genes had a vital impact on protein binding. Pathway analysis revealed that peroxisome was the most significantly enriched pathway, and m6A-modified genes were significantly enriched in the Wnt signaling pathway and calcium signaling pathway (Figure 3C).

Identification of differentially expressed genes

There were 12,869 genes detected at E19 and E27. The number of genes expressed in three individuals ranged from 11,226 to 11,494 at E19, while that ranged from 9,966 to 10,491 at E27. In general, the number of expressed genes was higher at E19 than that at E27. However, the overall gene expression levels were similar in both periods, which was shown from the gene expression box plot (Figure 4A). We also found 5,337 DEGs between these two stages, including 3,344 highly expressed genes at E27 and 1,993 highly expressed genes at E19 (Figures 4B–D).

Differentially expressed genes were selected for Gene Ontology (GO) and KEGG pathway enrichment analyses. According to the GO results, DEGs were not directly involved in the biological process, but more concentrated in the aspects of cellular components and molecular function. Eight hundred and three DEGs were enriched in the nucleus, and 681 DEGs were enriched in the cytoplasm. Furthermore, 723 DEGs participated in the molecular function of protein binding (Figure 4E). Through GO enrichment analysis, we found that the plasma membrane was the most significant term with the largest number of genes (436 genes). In addition, the selected DEGs were enriched in some skeletal muscle

TABLE 2 Descriptive statistics of sequenced data.

Sample_ID	Raw_Reads	Raw_Bases	Clean reads	Clean bases	Clean %	Q30%	GC%
E19_1_IP	87574914	13.14G	81524514	11.42G	86.90	93.80	56.63
E19_2_IP	84509772	12.68G	79209738	10.66G	84.07	94.38	55.98
E19_3_IP	74910974	11.24G	70367178	9.77G	86.99	93.59	55.31
E27_1_IP	71935808	10.79G	68192750	9.47G	87.75	94.02	55.20
E27_2_IP	74993630	11.25G	70405880	9.40G	83.54	93.80	54.21
E27_3_IP	75376878	11.31G	70574208	9.60G	84.93	94.30	55.61
E19_1_input	70475228	10.57G	65476148	8.59G	81.23	94.65	57.09
E19_2_input	88583582	13.29G	82522530	10.06G	75.72	94.86	56.84
E19_3_input	70391384	10.56G	66478036	9.16G	86.75	94.40	56.41
E27_1_input	86824864	13.02G	81526132	10.73G	82.36	94.75	55.79
E27_2_input	93174910	13.98G	88875790	11.83G	84.68	94.66	55.35
E27_3_input	94960962	14.24G	88872294	11.39G	79.97	94.54	56.37

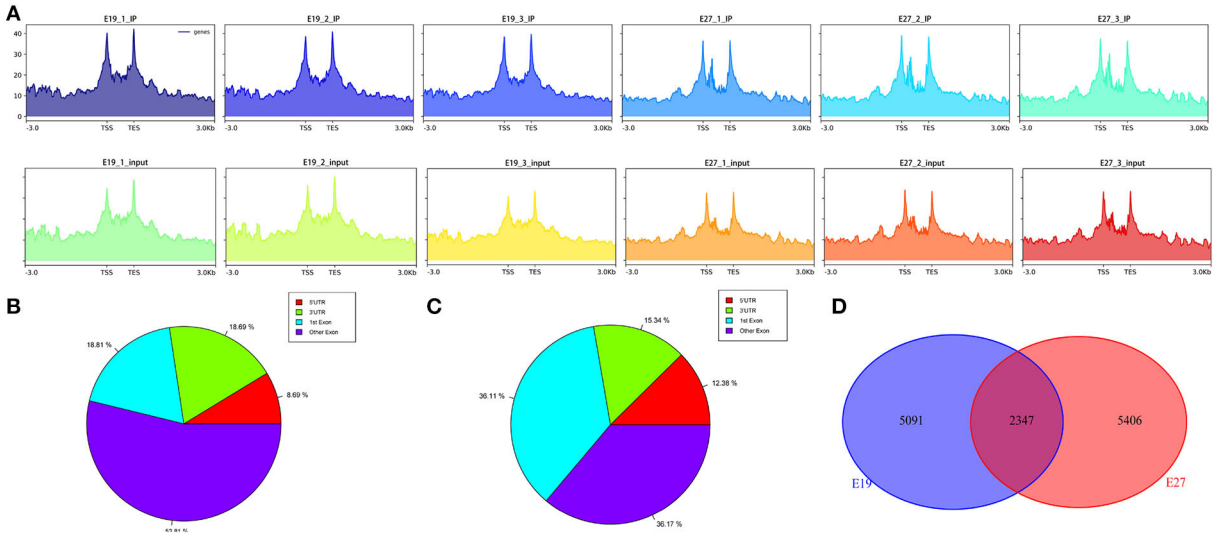


FIGURE 2
Analysis of transcriptome-wide m6A-seq data. **(A)** Distribution of reads in the upstream and downstream 3kb range from the TSS–TES (transcription end sites). The abscissa is the gene location, and the ordinate is the coverage depth of reads. **(B)** Pie chart of m6A peaks at E19. **(C)** Pie chart of m6A peaks at E27. **(D)** Pie chart of unique and shared m6A peaks between E19 and E27.

development and fat deposition GO terms, such as negative regulation of canonical Wnt signaling pathway, muscle organ development, activation of MAPKK activity, and activation of MAPK activity (Figure 4F). Moreover, KEGG analysis showed that DEGs were enriched in several pathways related to muscle development. For example, the Wnt signaling pathway was enriched, which is also related to skeletal muscle development. Additionally, pathways of dilated cardiomyopathy (DCM), fatty acid metabolism, viral myocarditis, and cardiac muscle contraction—which are associated with fat deposition and cardiac muscle development—were revealed through KEGG analysis (Figure 4G).

Conjoint analysis of m6A-seq and RNA-seq data at E19 and E27

As mentioned above, we had found 2,347 differential m6A peaks between the E19 and E27 stages, with 1,512 downregulated peaks and 835 upregulated peaks in abundance. Through conjoint analysis of m6A-seq and RNA-seq data, the downregulated peaks overlapped with 394 decreased expression genes (down–down genes) and 689 increased expression genes (down–up genes). Conversely, 227 genes with upregulation in m6A abundance showed downregulated gene expression (up–down genes) and 380 genes with upregulation in abundance

10.3389/fvets.2022.933850

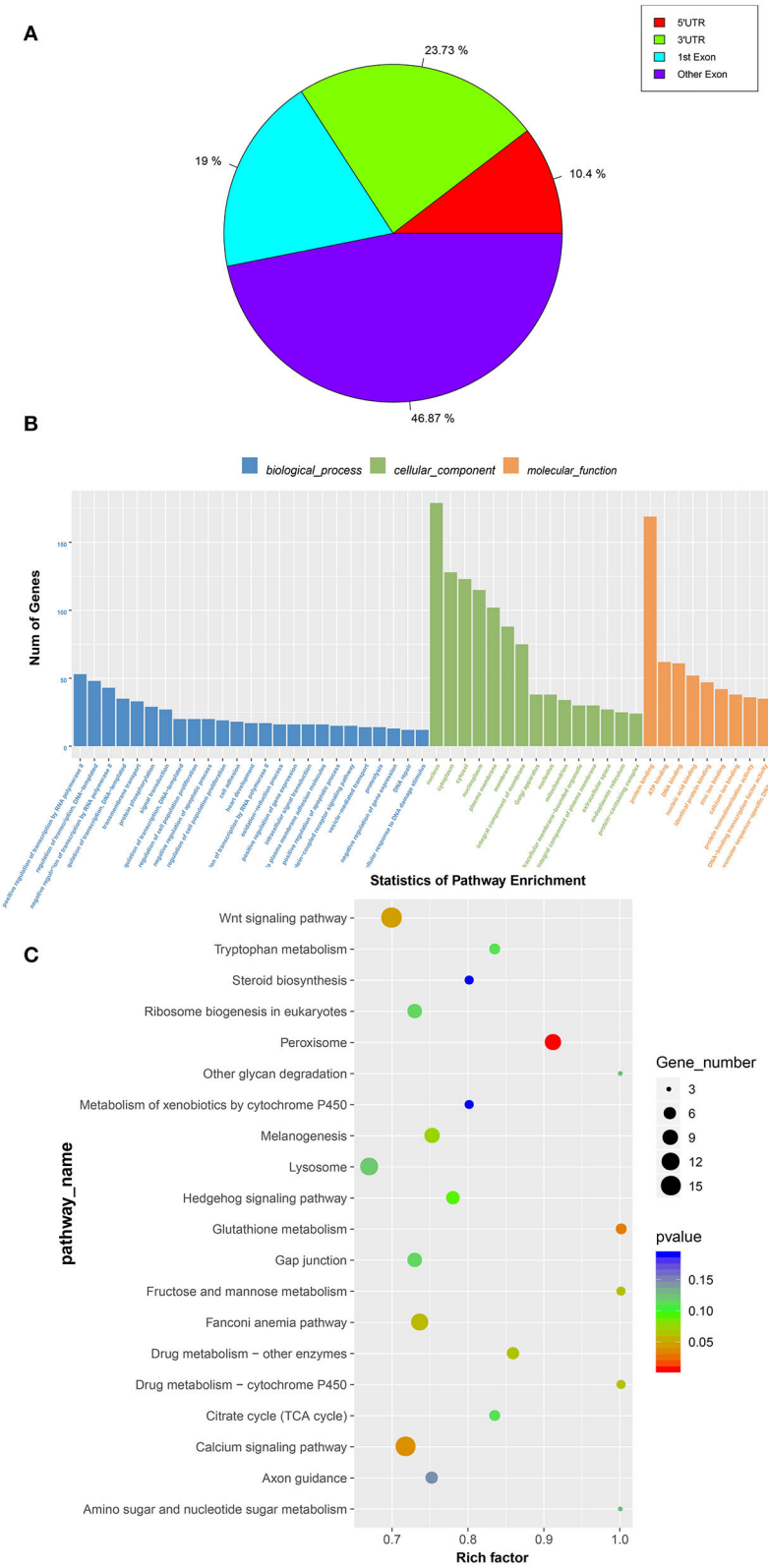


FIGURE 3
GO and KEGG analyses of m6A-modified genes. **(A)** Distribution of m6A-modified genes (MMGs) along genes. **(B)** Analysis of GO enrichment. **(C)** Statistics of KEGG pathway enrichment.

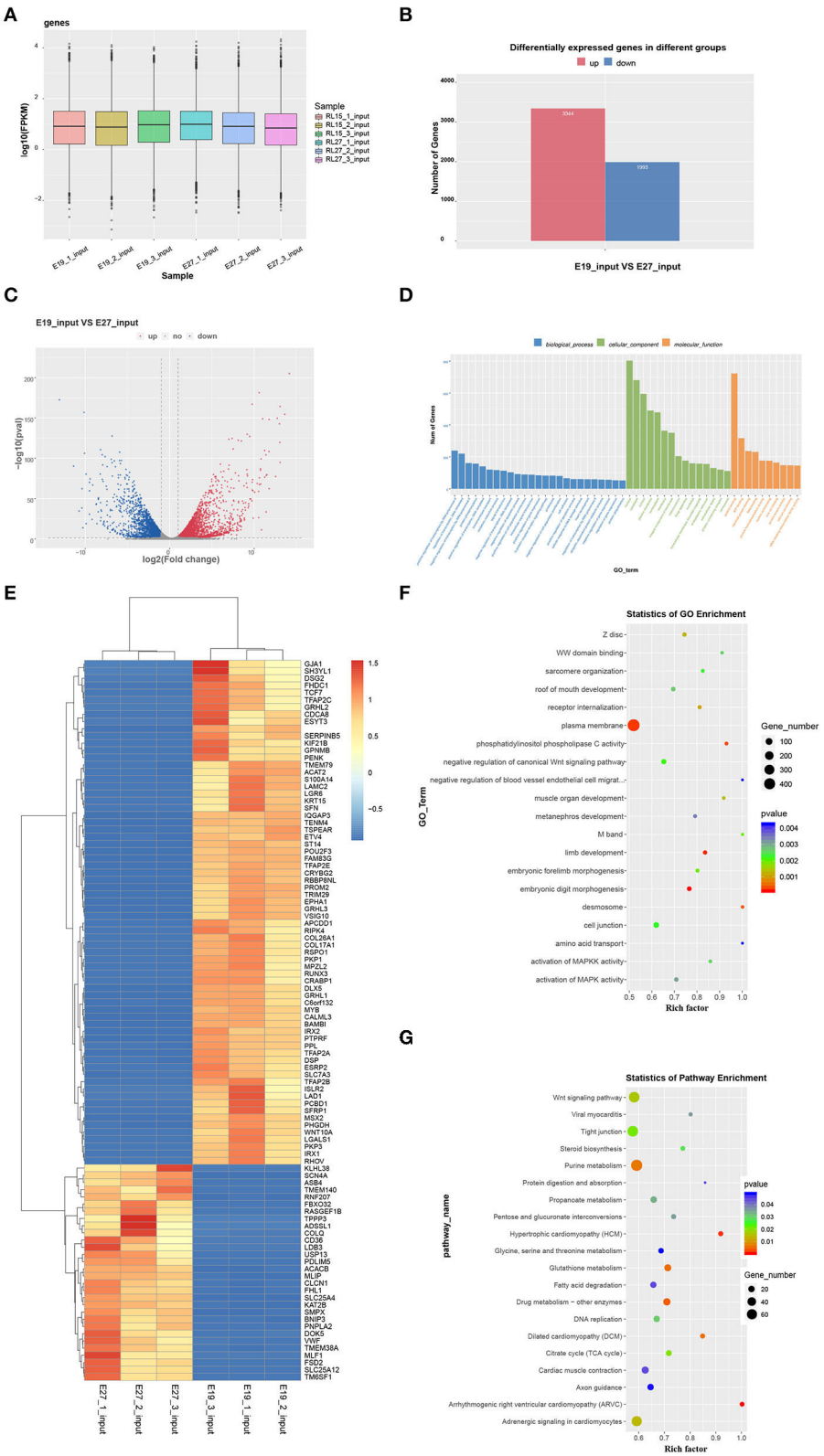


FIGURE 4
Global gene expression in the two stages and GO and KEGG analyses of DEGs. **(A)** Global expression levels of expressed genes. **(B)** Number of upregulated and downregulated DEGs between the two stages. **(C)** Volcanic map of DEGs. **(D)** Heatmap of DEGs. **(E)** Statistics of GO enrichment. **(F)** GO analysis of DEGs. **(G)** KEGG analysis of DEGs.

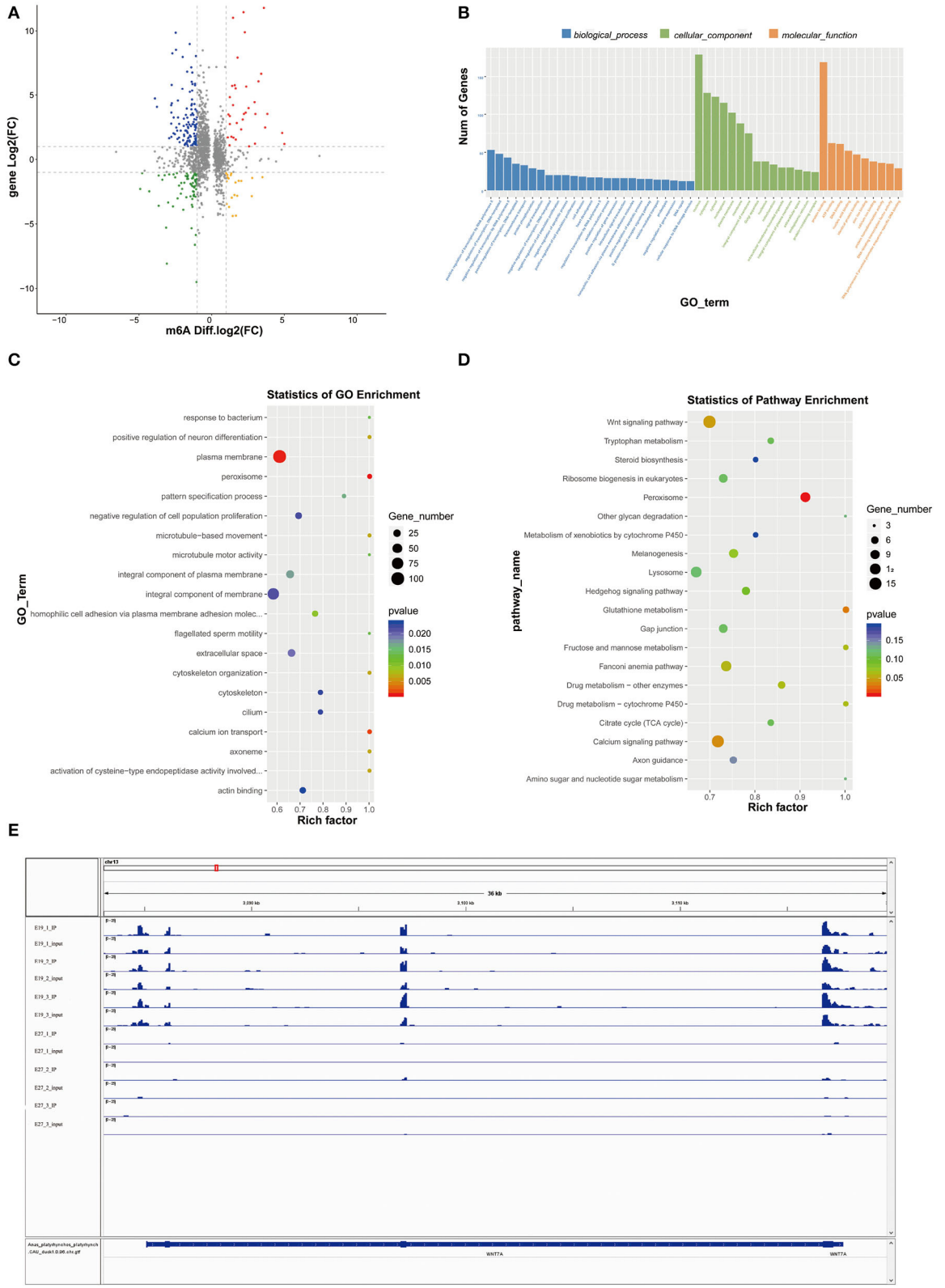


FIGURE 5
Conjoint analysis of m6A-seq and RNA-seq data. **(A)** Distribution of genes with a significant change in both m6A methylation level and gene expression between E19 and E27. Up−up, up−down represent genes with increased m6A methylation level and increased gene

(Continued)

FIGURE 5 (Continued)
expression and genes with increased m6A methylation level and decreased gene expression, respectively. Down–up, down–down represent genes with decreased m6A methylation level and increased gene expression and genes with decreased m6A methylation level and decreased gene expression, respectively. **(B)** GO terms of genes shared by MMGs and DEGs. **(C)** GO enrichment analysis of genes shared by MMGs and DEGs. **(D)** KEGG enrichment analysis of genes shared by MMGs and DEGs. **(E)** Visualization of m6A abundances in *WINT7a* mRNA transcripts at E19 and E27.

showed upregulated gene expression (up–up genes) (Figure 5A, Supplementary material 1).

Gene Ontology analysis for genes shared by MMGs and DEGs showed that plasma membrane, peroxisome, and calcium ion transport were the most significant three GO terms (Figures 5B,C), among which peroxisome is involved in β-oxidation of fatty acids and calcium ion transport is involved in muscle development), indicating some genes shared by MMGs and DEGs are potential regulators of skeletal muscle development and fat deposition.

Kyoto Encyclopedia of Genes and Genomes analysis showed that peroxisome (related to β-oxidation of fatty acids), Wnt signaling pathway, and calcium signaling pathway (tightly associated with skeletal muscle development) (Figure 5D) were the most significantly enriched pathways (the peroxisome pathway ranked the first), which were consistent with the GO results described above. In addition, we found many genes related to skeletal muscle development in MMGs such as *BCL9*, *SOX11*, *EPHB1*, *MYOCD*, *BVES*, *SLC8A3*, *ASB2*, *CFLAR*, *EPHB1*, *WNT7A*, and *SCN4A* (Table 3), suggesting that m6A modification plays crucial roles in duck muscle development. Among the skeletal muscle development-related genes, we picked out *Wnt7a* to compare the status of m6A modification levels in various stages and samples using Integrative Genomics Viewer (IGV) software (30). We also found that the m6A levels on the *Wnt7a* gene were significantly different when comparing E19 and E27 (Figure 5E).

Association analysis of miRNAs-seq, mRNA-seq, and m6A-seq data

We also tested the differentially expressed miRNAs (DEMs) between the E19 and E27 stages. There were 90 upregulated miRNAs and 102 downregulated miRNAs between the E19 and E27 stages. Through overlapping analysis of genes shared by MMGs and DEGs and the targets of DEMs, we identified six m6A-mRNA-regulated miRNAs, namely, cli-miR-1467-3p_1ss19AG, PC-3p-28816_21, efu-miR-9226_2ss4AG22GA, gga-miR-338-5p, gga-miR-338-3p, and apl-miR-11588-3p.

To verify the potential role of m6A-mRNA-regulated miRNAs, we performed the GO and KEGG analyses of the targets of m6A-mRNA-regulated miRNAs. The GO results showed that the most significant GO term ($P = 4.10E-04$) was phosphatidylinositol phosphorylation, which is involved in skeletal muscle development. Then, the second most significant GO term (phosphatidylinositol-mediated signaling) and other significant GO terms (kinase activity, phosphatidylinositol 3-kinase complex, and phosphatidylinositol 3-kinase complex, class IA) were all associated with skeletal muscle development (Figure 6A, Supplementary material 2). Being consistent with the GO results, the KEGG pathway analysis also found that some pathways were involved in skeletal muscle development, such as inositol phosphate metabolism, phosphatidylinositol signaling system, and focal adhesion (Figure 6B, Supplementary material 3).

Discussion

N6-methyladenosine is the most prevalent internal form of modification in polyadenylated mRNAs and long non-coding RNAs in higher eukaryotes, having been described in yeast, Arabidopsis, fruit flies, and mammals (31, 32). Recent studies have shown that m6A modifications to mRNA have a variety of biological functions and play key roles in gene expression regulation (33), animal development (16), and human diseases (34).

We aimed to describe the m6A modification profiles in embryonic breast muscle of ducks so as to lay a foundation for further exploring how m6A modifications contribute to the growth and development of duck breast muscle. In addition, we also did miRNAs-seq. MiRNA is also an important regulator of breast muscle development. The combination of the m6A-seq and miRNAs-seq will help us find the key target of breast muscle development.

Skeletal muscle development is a complex and multi-stage process that includes the formation of muscle cells into myotubes and the formation of muscle fibers (35, 36). However, almost no animals increase the number of muscle fibers after birth; thus, the amount of muscle meat production in adult livestock and poultry is determined during embryogenesis. Therefore, it is important to study embryonic muscle development (20).

Gu et al. (20) reported that E19 is the fastest point for breast muscle development. Before the E19 stage, breast muscle is mainly involved in muscle fiber proliferation events, while afterward, the crucial event is muscle fiber fusion to form more multinucleated myotubes. Therefore, we selected breast muscles

10.3389/fvets.2022.933850

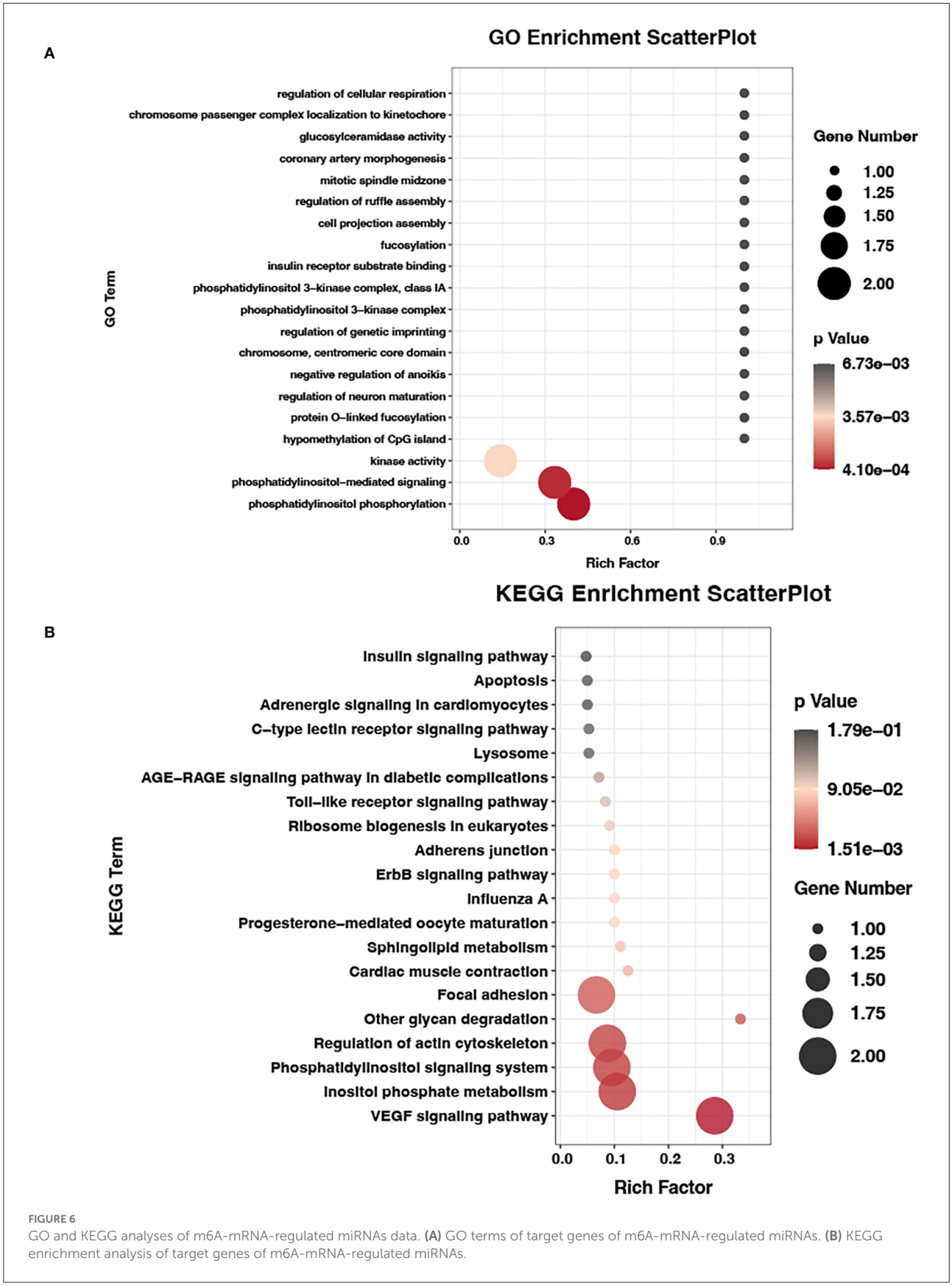


FIGURE 6
GO and KEGG analyses of m6A-mRNA-regulated miRNAs data. **(A)** GO terms of target genes of m6A-mRNA-regulated miRNAs. **(B)** KEGG enrichment analysis of target genes of m6A-mRNA-regulated miRNAs.

TABLE 3 List of 11 genes that exhibit a significant change in both m6A modification and mRNA expression related to muscle development in duck embryonic breast muscle tissues.

Gene name	Chromosome	Peak start	Peak end	Peak length	Transcript ID	Pattern	*P*-value
BCL9	chr1	95237263	95242335	5073	ENSAPLT00000032721	Down–up	1.79414E-11
MYC	chr2	146787275	146787752	478	ENSAPLT00000014574	Down–up	0.000850783
BVES	chr3	71742456	71742814	359	ENSAPLT00000040909	Up–down	9.19716E-45
SOX11	chr3	78759336	78759927	592	ENSAPLT00000046689	Down–up	6.90452E-07
SLC8A3	chr5	28282453	28282632	180	ENSAPLT00000016237	Up–down	5.18455E-16
ASB2	chr5	48558337	48559238	902	ENSAPLT00000011413	Up–down	9.51065E-58
CFLAR	chr7	98022	99193	1172	ENSAPLT00000015969	Up–down	0.045676673
EPHB1	chr9	10474372	10474638	267	ENSAPLT00000034231	Up–down	1.85369E-09
WNT7A	chr13	3117467	3117616	150	ENSAPLT00000014453	Up–down	0.000245641
MYOCD	chr19	620339	620964	626	ENSAPLT00000046004	Up–down	1.04238E-21
SCN4A	chr28	715525	716182	658	ENSAPLT00000005205	Up–down	5.4444E-155

at the E13, E19, and E27 stages for the preliminary experiment and found that the expression levels of m6A and methylation modification enzymes were all highest at E19. This is consistent with the duck breast muscle development model and suggests that m6A methylation modification plays some key role in the development of embryonic breast muscle of ducks. These results encouraged us to perform m6A sequencing of embryonic breast muscle at the E19 and E27 stages.

Through m6A-seq, we obtained a list of m6A peaks that overlapped with genes at the E19 (7,438) and E27 (7,753) stages, indicating that m6A modification might be a common approach for duck gene regulation. The methylated genes obtained in this study are much higher than those detected in pigs and chickens, respectively, finding 5,864 and 3,303 methylated genes for muscle tissues and adipose tissues of pigs, and 2,893 and 4,593 transcripts for pre- and post-selection follicles (19, 37).

Recently, two independent studies combining m6A immunoprecipitation with high-throughput analysis revealed that the m6A modification tends to occur in the termination codon, 3′UTRs, mRNA exons, and protein-coding regions (12, 38). Our research also manifested similar distribution patterns in which the peaks from m6A-IP samples were highly accumulated around the other exons (except the first exon) in two embryonic periods, accounting for 53.81 and 36.17%, respectively (Figures 3B,C). At the E19 stages, peaks on the first exon and 3′UTR accounted for 18.81 and 18.69%, respectively, while peaks on the first exon and the other exons both accounted for ∼36%, suggesting the conservation of m6A distribution in various species.

The skeletal muscle development and fat deposition are both complicated multi-step processes involving some crucial signaling pathways, e.g., initiation and mediation. For skeletal muscle development (39), the Wnt signaling pathway (40), the activation of MAPK activity (41), and the calcium signaling pathway (42) are all tightly associated with skeletal muscle development. For fat deposition, both fatty acid metabolism and peroxisome (43) play key roles. In this study, many MMGs were enriched in both muscle-related pathways (Wnt signaling pathway, calcium signaling pathway, and/or MAPK activity) and fat-related pathways (peroxisome). m6A modification was tightly related to biological processes, including skeletal development and fat deposition, which suggested m6A modification might play crucial roles in duck breast muscle development and fat deposition. Moreover, a negative correlation between m6A methylation enrichment and gene expression levels was found in chicken follicles (37). In addition, our previous study also indicated that E19 was the fastest point of breast muscle development (20). Therefore, we propose that MMGs with lower m6A levels may be the positive regulators for the breast muscle development of ducks, while MMGs with higher m6A levels might be the negative regulators. However, the results obtained above need to be validated in future by molecular experiments.

N6-methyladenosine modification has a regulatory effect on the mRNA of the gene, thus affecting the gene's function. "Writers" were responsible for determining the sex of Drosophila development by adding m6A modifications to the pre-mRNA of Sxl (44). YTHDF2 can regulate the mRNA level during the maternal-to-zygotic transition (MZT) and regulate the development of zebrafish offspring (45). Therefore, association analysis of MMGs and DEGs is important in investigating the regulatory roles of m6A.

In this study, we overlapped MMGs and DEGs and obtained the shared genes. Thus, the shared genes might be the genes affected by m6A. We further picked out 11 muscle-related development genes from the shared genes, including four "up–up" genes (*BCL9*, *SOX11*, *EPHB1*, and *MYOCD*) and seven "down–down" genes (*BVES, SLC8A3, ASB2*, *CFLAR*, *EPHB1*,

Wnt1a, and *SCN4A*) (Table 3), which can be used to explore the regulation mechanism of m6A modification in embryonic breast muscle of ducks. In particular, *Wnt7a* is implicated in playing roles in homeostasis maintenance of skeletal muscle, and *Wnt7a* treatment may be potentially applied in skeletal muscle dystrophy (46). In addition, *Wnt7a* induces satellite cell expansion, myofiber hyperplasia, and hypertrophy in rat craniofacial muscle (47). Therefore, we selected *Wnt7a* for visualization analysis.

It has been reported that the transcript of argonaute-2 (AGO2), a catalytic protein in miRNA-mediated gene silencing, was highly methylated in young peripheral blood mononuclear cells although less so following aging (48). The study also showed a correlation of m6A-methylated AGO2 mRNA with miRNA expression, and it indicated a negative effect of m6A methylation on miRNA expressions during cellular senescence. In this study, we found that six DEMs with their target genes overlapped with genes shared by DEGs and DMGs, suggesting that these six DEMs are regulated by m6A modification. The GO and KEGG analyses for the targets of the six DEMs showed many significantly enriched GO terms or KEGG pathways involved in the regulation of skeletal muscle development. Among them, phosphatidylinositol 3-kinase complex and phosphatidylinositol 3-kinase complex class IA may regulate AMPK activity (49) and subsequently promote skeletal muscle regeneration (50). In addition, phosphatidylinositol phosphorylation, phosphatidylinositol-mediated signaling, and inositol phosphate metabolism are all related to the phosphatidylinositol signaling system. Safi et al. (51) showed that the PI3K pathway, a pathway belonging to the phosphatidylinositol signaling system, can impede the effect of CKIP-1 on C2C12 cell differentiation. Based on our results and the available literature data, we speculate that m6A modification might play a key role in the skeletal muscle development of ducks by affecting miRNA.

In conclusion, we compared the homologous methylase sequences between ducks and humans and illustrated the overall m6A level and the expression of methylases at the E19 and E27 stages. We also revealed the global m6A modification patterns in duck embryonic breast muscles and found that MMGs were enriched in skeletal muscle development-related pathways. In addition, our results strongly indicated that genes shared by DEGs and MMGs might be associated with skeletal muscle development and fat deposition. Finally, we found miRNAs were also regulated by m6A modification, revealed by association analysis of miRNA-seq, RNA-seq, and m6A-seq data.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The animal study was reviewed and approved by Animal Care and Use of Chinese Academy of Tropical Agricultural Sciences.

Author contributions

LG and TX prepared the manuscript and designed the experiments. LG, SZ, BL, QJ, TX, YH, DL, MX, LH, FW, XZ, ZC, and WS collected the samples. TX was responsible for the design and direction of the experiment. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China General Program (31972553), Key Research and Development Programs of Hainan Province (ZDYF2022XDNY154), Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (630032017034 and 1630032022013), Chinese Modern Technology System of Agricultural Industry (CARS-42-50), and Special Funds for Central Government Guiding Local Science and Technology Development (ZY2019HN01).

Acknowledgments

The authors appreciate Yunsheng Zhang for the sample selection and Fei Qiao for the modification of the initial manuscript.

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 SR declared a shared affiliation with the authors YH and SZ to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.933850/full#supplementary-material

References

1. Boulias K, Toczydlowska-Socha D, Hawley B, Liberman N, Takashima K, Zaccara S, et al. Identification of the m(6)Am methyltransferase PCIF1 reveals the location and functions of m(6)Am in the transcriptome. *Mol Cell.* (2019) 75:631. doi: 10.1016/j.molcel.2019.06.006

2. Meyer KD, Jaffrey SR. Rethinking m(6)A readers, writers, and erasers. *Annu Rev Cell Dev Biol.* (2017) 33:319–42. doi: 10.1146/annurev-cellbio-100616-060758

3. Chang R, Huang Z, Zhao S, Zou J, Li Y, Tan S. Emerging roles of FTO in neuropsychiatric disorders. *Biomed Res Int.* (2022) 2022:2677312. doi: 10.1155/2022/2677312

4. Li H, Zhong Y, Cao G, Shi H, Liu Y, Li L, et al. METTL3 promotes cell cycle progression via m(6)A/YTHDF1-dependent regulation of CDC25B translation. *Int J Biol Sci.* (2022) 18:3223–36. doi: 10.7150/ijbs.70335

5. Shen H, Luo B, Wang Y, Li J, Hu Z, Xie Q, et al. Genome-Wide Identification, Classification and Expression Analysis of m(6)A Gene Family in Solanum lycopersicum. *Int J Mol Sci.* (2022) 23:4522. doi: 10.3390/ijms23094522

6. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A.* (1974) 71:3971–5. doi: 10.1073/pnas.71.10.3971

7. Levis R, Penman S. 5'-terminal structures of poly(A)+ cytoplasmic messenger RNA and of poly(A)+ and poly(A)- heterogeneous nuclear RNA of cells of the dipteran Drosophila melanogaster. *J Mol Biol.* (1978) 120:487–515. doi: 10.1016/0022-2836(78)90350-9

8. Narayan P, Ayers DF, Rottman FM, Maroney PA, Nilsen TW. Unequal distribution of N6-methyladenosine in influenza virus mRNAs. *Mol Cell Biol.* (1987) 7:1572–5. doi: 10.1128/MCB.7.4.1572

9. Pan T. N6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends Biochem Sci.* (2013) 38:204–9. doi: 10.1016/j.tibs.2012.12.006

10. Shi HL, Wei JB, He C. Where, When, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol Cell.* (2019) 74:640–50. doi: 10.1016/j.molcel.2019.04.025

11. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol.* (2014) 10:93–5. doi: 10.1038/nchembio.1432

12. Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* (2014) 24:177–89. doi: 10.1038/cr.2014.3

13. Shen DD, Suo FZ, Song QM, Chang J, Zhang T, Hong JJ, et al. Development of formaldehyde dehydrogenase-coupled assay and antibody-based assays for ALKBH5 activity evaluation. *J Pharm Biomed Anal.* (2019) 162:9–15. doi: 10.1016/j.jpba.2018.09.018

14. Mathiyalagan P, Adamiak M, Mayourian J, Sassi Y, Liang Y, Agarwal N, et al. FTO-Dependent N(6)-Methyladenosine Regulates Cardiac Function During Remodeling and Repair. *Circulation.* (2019) 139:518–32. doi: 10.1161/CIRCULATIONAHA.118.033794

15. Mauer J, Jaffrey SR. FTO, m(6) Am, and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Lett.* (2018) 592:2012–22. doi: 10.1002/1873-3468.13092

16. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science.* (2018) 361:1346–9. doi: 10.1126/science.aau1646

17. Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.* (2018) 28:616–24. doi: 10.1038/s41422-018-0040-8

18. Zhao T, Zhao R, Yi X, Cai R, Pang W. METTL3 promotes proliferation and myogenic differentiation through m(6)A RNA methylation/YTHDF1/2 signaling axis in myoblasts. *Life Sci.* (2022) 298:120496. doi: 10.1016/j.lfs.2022.120496

19. Tao X, Chen J, Jiang Y, Wei Y, Chen Y, Xu H, et al. Transcriptome-wide N (6) -methyladenosine methylome profiling of porcine muscle and adipose tissues reveals a potential mechanism for transcriptional regulation and differential methylation pattern. *BMC Genomics.* (2017) 18:336. doi: 10.1186/s12864-017-3719-1

20. Gu LH, Xu TS, Huang W, Xie M, Shi WB, Sun SD, et al. Developmental characteristics of pectoralis muscle in Pekin duck embryos. *Genet Mol Res.* (2013) 12:6733–42. doi: 10.4238/2013.December.13.6

21. Xu TS, Gu LH, Huang W, Xia WL, Zhang YS, Zhang YG, et al. Gene expression profiling in Pekin duck embryonic breast muscle. *PLoS ONE.* (2017) 12:e0174612. doi: 10.1371/journal.pone.0174612

22. Gu L, Xu T, Huang W, Xie M, Sun S, Hou S. Identification and profiling of microRNAs in the embryonic breast muscle of pekin duck. *PLoS ONE.* (2014) 9:e86150. doi: 10.1371/journal.pone.0086150

23. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* (2015) 12:357–60. doi: 10.1038/nmeth.3317

24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

25. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* (2015) 4:5005. doi: 10.7554/eLife.05005

26. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* (2009) 19:92–105. doi: 10.1101/gr.082701.108

27. Nam JW, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, et al. Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol Cell.* (2014) 53:1031–43. doi: 10.1016/j.molcel.2014.02.013

28. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* (2008) 36:D149–153. doi: 10.1093/nar/gkm995

29. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. *Genome Biol.* (2003) 5:R1. doi: 10.1186/gb-2003-5-1-r1

30. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* (2013) 14:178–92. doi: 10.1093/bib/bbs017

31. Krug RM, Morgan MA, Shatkin AJ. Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *J Virol.* (1976) 20:45–53. doi: 10.1128/jvi.20.1.45-53.1976

32. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci U S A.* (1984) 81:5667–71. doi: 10.1073/pnas.81.18.5667

33. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell.* (2017) 169:1187–200. doi: 10.1016/j.cell.2017.05.045

34. Hsu PJ, Shi H, He C. Epitranscriptomic influences on development and disease. *Genome Biol.* (2017) 18:197. doi: 10.1186/s13059-017-1336-6

35. Patel K, Christ B, Stockdale FE. Control of muscle size during embryonic, fetal, and adult life. *Results Probl Cell Differ.* (2002) 38:163–86. doi: 10.1007/978-3-540-45686-5_8

36. Buckingham M. Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr Opin Genet Dev.* (2006) 16:525–32. doi: 10.1016/j.gde.2006.08.008

37. Fan Y, Zhang C, Zhu G. Profiling of RNA N6-methyladenosine methylation during follicle selection in chicken ovary. *Poult Sci.* (2019) 98:6117–24. doi: 10.3382/ps/pez277

38. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell.* (2012) 149:1635–46. doi: 10.1016/j.cell.2012.05.003

39. Roffe S, Hagai Y, Pines M, Halevy O. Halofuginone inhibits Smad3 phosphorylation via the PI3K/Akt and MAPK/ERK pathways in muscle cells: effect on myotube fusion. *Exp Cell Res.* (2010) 316:1061–9. doi: 10.1016/j.yexcr.2010.01.003

40. Girardi F, Le Grand F. Wnt signaling in skeletal muscle development and regeneration. *Prog Mol Biol Transl Sci.* (2018) 153:157–79. doi: 10.1016/bs.pmbts.2017.11.026

41. Tomida T, Adachi-Akahane S. [Roles of p38 MAPK signaling in the skeletal muscle formation, regeneration, and pathology]. *Nihon Yakurigaku Zasshi.* (2020) 155:241–7. doi: 10.1254/fpj20030

42. Kuo IY, Ehrlich BE. Signaling in muscle contraction. *Cold Spring Harb Perspect Biol.* (2015) 7:a006023. doi: 10.1101/cshperspect.a006023

43. Baek JH, Kim DH, Lee J, Kim SJ, Chun KH. Galectin-1 accelerates high-fat diet-induced obesity by activation of peroxisome proliferator-activated receptor gamma (PPARgamma) in mice. *Cell Death Dis.* (2021) 12:66. doi: 10.1038/s41419-020-03367-z

44. Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, et al. m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature.* (2016) 540:301–4. doi: 10.1038/nature20577

45. Zhao BS, Wang X, Beadell AV, Lu Z, Shi H, Kuuspalu A, et al. m(6)A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature.* (2017) 542:475–8. doi: 10.1038/nature21355

46. Lan L, Wang W, Huang Y, Bu X, Zhao C. Roles of Wnt7a in embryo development, tissue homeostasis, and human diseases. *J Cell Biochem.* (2019) 120:18588–98. doi: 10.1002/jcb.29217

47. Cheng X, Huang H, Luo X, Shi B, Li J. Wnt7a induces satellite cell expansion, myofiber hyperplasia and hypertrophy in rat craniofacial muscle. *Sci Rep.* (2018) 8:10613. doi: 10.1038/s41598-018-28917-6

48. Li Z, Zhao P, Xia Q. Epigenetic Methylations on N6-Adenine and N6-Adenosine with the same Input but Different Output. *Int J Mol Sci.* (2019) 20:2931. doi: 10.3390/ijms20122931

49. Liu Y, Nguyen PT, Wang X, Zhao Y, Meacham CE, Zou Z, et al. TLR9 and beclin 1 crosstalk regulates muscle AMPK activation in exercise. *Nature.* (2020) 578:605–9. doi: 10.1038/s41586-020-1992-7

50. Mounier R, Theret M, Lantier L, Foretz M, Viollet B. Expanding roles for AMPK in skeletal muscle plasticity. *Trends Endocrinol Metab.* (2015) 26:275–86. doi: 10.1016/j.tem.2015.02.009

51. Safi A, Vandromme M, Caussanel S, Valdacci L, Baas D, Vidal M, et al. Role for the pleckstrin homology domain-containing protein CKIP-1 in phosphatidylinositol 3-kinase-regulated muscle differentiation. *Mol Cell Biol.* (2004) 24:1245–55. doi: 10.1128/MCB.24.3.1245-1255.2004

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 04 November 2022
DOI 10.3389/fvets.2022.954882

OPEN ACCESS

EDITED BY
Ana Fabrícia Braga Magalhães,
Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

REVIEWED BY
Weihao Chen,
Yangzhou University, China
Wenping Hu,
Institute of Animal Sciences (CAAS), China

*CORRESPONDENCE
Gerelt Borjigin
bor_gerelt07@imau.edu.cn

SPECIALTY SECTION
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

RECEIVED 27 May 2022
ACCEPTED 20 October 2022
PUBLISHED 04 November 2022

CITATION
He X, Wu R, Yun Y, Qin X, Huang Y, Chen L, Han Y, Wu J, Sha L and Borjigin G (2022) MicroRNA and circular RNA profiling in the deposited fat tissue of Sunite sheep. *Front. Vet. Sci.* 9:954882. doi: 10.3389/fvets.2022.954882

MicroRNA and circular RNA profiling in the deposited fat tissue of Sunite sheep

Xige He[1], Rihan Wu[2], Yueying Yun[1,3], Xia Qin[1], Yajuan Huang[1], Lu Chen[1], Yunfei Han[1], Jindi Wu[1], Lina Sha[1] and Gerelt Borjigin[1]*

[1]College of Food Science and Engineering, Inner Mongolia Agricultural University, Hohhot, China, [2]College of Biochemistry and Engineering, Hohhot Vocational College, Hohhot, China, [3]School of Life Science and Technology, Inner Mongolia University of Science and Technology, Baotou, China

As the most typical deposited fat, tail fat is an important energy reservoir for sheep adapted to harsh environments and plays an important role as a raw material in daily life. However, the regulatory mechanisms of microRNA (miRNA) and circular RNA (circRNA) in tail fat development remain unclear. In this study, we characterized the miRNA and circRNA expression profiles in the tail fat of sheep at the ages of 6, 18, and 30 months. We identified 219 differentially expressed (DE) miRNAs (including 12 novel miRNAs), which exhibited a major tendency to be downregulated, and 198 DE circRNAs, which exhibited a tendency to be upregulated. Target gene prediction analysis was performed for the DE miRNAs. Functional analysis revealed that their target genes were mainly involved in cellular interactions, while the host genes of DE circRNAs were implicated in lipid and fatty acid metabolism. Subsequently, we established a competing endogenous RNA (ceRNA) network based on the negative regulatory relationship between miRNAs and target genes. The network revealed that upregulated miRNAs play a leading role in the development of tail fat. Finally, the ceRNA relationship network with oar-miR-27a_R-1 and oar-miR-29a as the core was validated, suggesting possible involvement of these interactions in tail fat development. In summary, DE miRNAs were negatively correlated with DE circRNAs during sheep tail fat development. The multiple ceRNA regulatory network dominated by upregulated DE miRNAs may play a key role in this developmental process.

KEYWORDS

Sunite sheep, tail fat, microRNA, circular RNA, competing endogenous RNA

Introduction

Adipose tissue is distributed in various parts of the sheep body and plays a crucial role in maintaining the balance of homeostatic metabolic processes in the body. Generally, adipose tissue can be found in the subcutaneous layer under the skin, around the kidneys, and within the abdominal cavity, and the tail, especially the tail fat is one of the most typical deposited fat. The "fat tail" trait of sheep is regarded as an adaptive response to the harsh environments and the fat stored in the tail is a valuable reserve for sheep during migration and in winter when food is scarce (1).

MicroRNAs (miRNAs) are single-stranded non-coding RNA molecules which are encoded by genes and bind to target mRNA transcripts *via* complementary base-pairing to exert their effects on expression (2). Several recent studies have analyzed sheep tail fat *via* miRNA-seq. The miRNA expression profiles of tail fat from Guangling large-tailed sheep and small-tailed Han sheep have been analyzed (2). A total of 40 differentially expressed (DE) conserved miRNAs were identified, in addition to 150 significantly expressed miRNAs, suggesting that these may play a role in the regulation of tail fat metabolism. Another study compared tail fat miRNA expression between Kazakhstan sheep (fat-tailed) and Tibetan sheep (thin-tailed) (3), revealing 539 miRNAs that were found in both breeds, of which 35 were novel. These miRNAs were involved in the MAPK, FoxO, and Wnt signaling pathways *via* their target mRNAs, thus influencing fat deposition and lipid metabolism in the fat tail. Current miRNA-seq studies have focused on other sheep tissues such as intramuscular fat (4), testis (5), and uterus (6), but studies related to tail adipose tissue are relatively lacking. Thus, the involvement of miRNAs in sheep tail fat metabolism requires further in-depth research. Different types of non-coding RNA perform a wide variety of biological functions and are involved in the regulation of diverse important pathways. These constitute endogenous RNA (ceRNA) networks, in which miRNAs play a central role. Recent studies have found that circular RNA (circRNAs) act as miRNA sponges, isolating miRNAs through competitive interactions with target mRNAs (7). CircRNAs are another class of non-coding RNAs, which are formed by covalent binding (reverse splicing) of the 5′ end of a linear RNA to the 3′ end (8). Since circRNAs have no free 5′ or 3′ end, they are not cleaved by exonucleases, which makes them more stable than most linear RNAs (9). A recent study in pig adipose tissue revealed that circRNA26852 and circRNA11897 target genes may be involved in adipocyte differentiation and lipid metabolism (10). Similarly, in buffalo, circRNAs 19:45387150|45389986 and 21:6969877|69753491 were shown to regulate fat deposition (11). In our previous study, we characterized the lncRNA and mRNA expression profiles of tail adipose tissue from Sunite sheep (SS) (12). The results showed that a total of 223 differentially expressed genes (DEGs) and 148 differentially expressed lncRNAs were found in tail fat, and the interaction of these genes may be involved in the metabolism of sheep tail fat. However, related research on miRNA and circRNA in SS tail fat is still lacking, including the regulation mechanism of fat metabolism.

SS is a Mongolian meat breed, and the distinctive phenotypic feature of SS is the fat tail. SS is a representative local superior breed within Inner Mongolia, with cold tolerance, drought resistance, rapid growth and development, high vitality, delicate flesh, and good flavor, which altogether make them very popular among consumers. The fat tail, of these sheep is the most typical deposited fat and helps SS adapt to harsh conditions such as cold, drought, and food shortages (13, 14), making them more adaptable than other breeds (15). The tail fat increases continuously with age, reaching a weight of ∼3–4.5 kg at 30 months of age. As a by-product of mutton, tail fat is widely used as a raw material for the production of various daily necessities. Also, it can be an important source of dietary fat (1, 16), providing the human body with the energy it needs. While the deposition of tail fat in sheep may affect their intramuscular fat content to a certain extent (17), tail fat metabolic regulation is yet to be investigated. SS are mainly raised naturally grazing in the Sunite grassland of Inner Mongolia. These sheep are accustomed to autonomous activity and completely voluntary feeding. Thus, the deposition and metabolism of SS tail fat might be affected by a large number of potential factors (including feeding behavior, aging processes, pasture changes, seasonal alterations, and climate impact). In the present work, the expression profiles of miRNA and circRNA in deposited fat (tail fat) from SS at the ages of 6 months (6 M), 18 months (18 M), and 30 months (30 M) were analyzed. We sought to elucidate mechanisms underlying tail fat metabolism by constructing a ceRNA co-regulatory network, thus providing valuable insights into the transcriptome associated with sheep fat tissue metabolism and utilization of by-products of meat breeds of sheep.

Materials and methods

Sample collection

The experimental animals were nine castrated Sunite rams at 6 (6 M, $n = 3$), 18 (18 M, $n = 3$), and 30 months of age (30 M, $n = 3$). All sheep were raised under the same conditions (food, water source, and environment) in the Xilingol grassland. After slaughter (in October), adipose tissue from the tail fat (top 1/3) from each sheep was obtained and immediately frozen in liquid nitrogen.

RNA extraction, sequencing, and transcript assembly

Total RNA was extracted from each sample using TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. The quantity and purity of total RNA were determined using a Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA), respectively. Approximately 1 μg of total RNA was used for small RNA library construction with TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, USA), and single-end sequencing (36 or 50 bp) was performed on an Illumina HiSeq 2500. Subsequently, raw reads were subjected to ACGT101-miR (LC Sciences, Houston, TX, USA) to remove repeats, junk and low complexity, adapter dimers, and common RNA families (rRNA, tRNA, snRNA, and snoRNA).

Then, unique sequences with lengths of 18–26 nucleotides were mapped to specific species precursors and the genome in miRBase 21.0 *via* BLAST search in order to identify known and novel miRNAs.

Approximately 10 μg of total RNA was used to deplete ribosomal RNA using the Epicenter Ribo-Zero Gold Kit (Illumina, San Diego, USA) as per manufacturer instructions. The remaining RNA fragments were then reverse-transcribed to form the final complementary DNA (cDNA) library using an RNA-seq Library Preparation Kit (Illumina, San Diego, USA) according to the manufacturer's protocol. Finally, paired-end sequencing on an Illumina HiSeq 4000 was performed following the manufacturer's protocol. Mapped reads were assembled into circRNA using CIRCExplorer (18, 19). Tophat-fusion and CIRCExporer were used to identify the back-splicing reads among the unmapped reads.

Differential expression analysis

The differential expression of miRNAs based on normalized deep-sequencing counts was analyzed using a *T*-test. The threshold for differential expression was set at $P < 0.05$. SRPBM was used to normalize the expression of circRNA in our study (20). Differentially expressed circRNAs were selected based on $|\log2 \text{ (fold change)}| > 1$ and $P < 0.05$ using the R package Ballgown (21). Further, we performed trend enrichment analysis on the DE miRNAs using the Short Time-series Expression Miner (STEM) software (22) as well as significant enrichment analysis with a threshold of $P < 0.05$.

Target gene prediction and functional analysis

To explore the functions of DE miRNAs, circRNAs and mRNAs were predicted as miRNA targets using TargetScan (23) and miRanda (24). A TargetScan Score > 50 and miRanda Energy < −20 were considered indicative of a targeting relationship. We then performed GO and KEGG analyses of DE miRNA targets and the host genes of DE circRNAs using in-house scripts. Statistical significance was set at $P < 0.05$. SS tail fat mRNA (including circRNA) and miRNA data were deposited in the NCBI Sequence Read Archive (SRA) database under accession numbers PRJNA791005 and PRJNA790717, respectively.

Construction of ceRNA co-expression network

After determining the target relationships of DE miRNAs using TargetScan and miRanda, we screened miRNA-targets with negative regulatory relationships (e.g., down regulation-up regulation) from these target relationships to build the mRNA-miRNA-circRNA co-expression networks. DE mRNAs were filtered based on $P < 0.05$ and $|\log2 \text{ (fold change)}| > 1$. Cytoscape (version 3.9.0) was used to visualize the network.

qRT-PCR validation

In our study, six DE miRNAs and eight DE circRNAs were randomly selected to validate RNA-seq data using real-time quantitative PCR. The expression of each miRNA and circRNA was calculated *via* the $2^{-\Delta\Delta CT}$ method, with GAPDH (25, 26) and U6 used as reference genes for circRNA and miRNA, respectively. The primer information is shown in the Supplementary Table S1.

Dual-luciferase gene reporter analysis

The psiCHECK2-target WT (wild type) gene was synthesized by inserting a target gene fragment containing the miRNA-binding sequence into the luciferase gene of the psiCHECK-2 vector. The mutant vector pCK TCP1-M was created by mutating the miRNA-binding sites using overlapping extension PCR. HEK293T cells were seeded into 24-well plates at a density of 1×10^5 cells/well and incubated at 37°C overnight. The miRNA mimics, psiCHECK2-target WT gene, and psiCHECK2-target MUT (mutant traits) gene were transfected into cells. At 48 h post-transfection, the Renilla luciferase activity/firefly luciferase activity was determined using the dual-luciferase reporter gene assay system (Promega).

Results

Summary of RNA-seq analysis

We obtained 1,942 miRNAs, including 392 novel miRNAs, and the majority of reads were ∼21–23 nucleotides (nt) in length, which corresponds to the typical length following Dicer enzyme cleavage (Figure 1A). We also identified 17,531 circRNAs and 4,767 host genes. As shown in Figure 1B, 39.75% of the circRNAs were transcribed from a unique mRNA. UTRN (0.02%) was the most common host gene, giving rise to 109 circRNAs. Researchers have shown that the UTRN gene is related to pig intramuscular fat (27). In addition, ACACA, which plays a key role in the regulation of fatty acid synthesis, was a host gene of 60 circRNAs. Thus, these circRNAs may exert a potential regulatory effect on fat metabolism in sheep tails by modulating host gene expression.

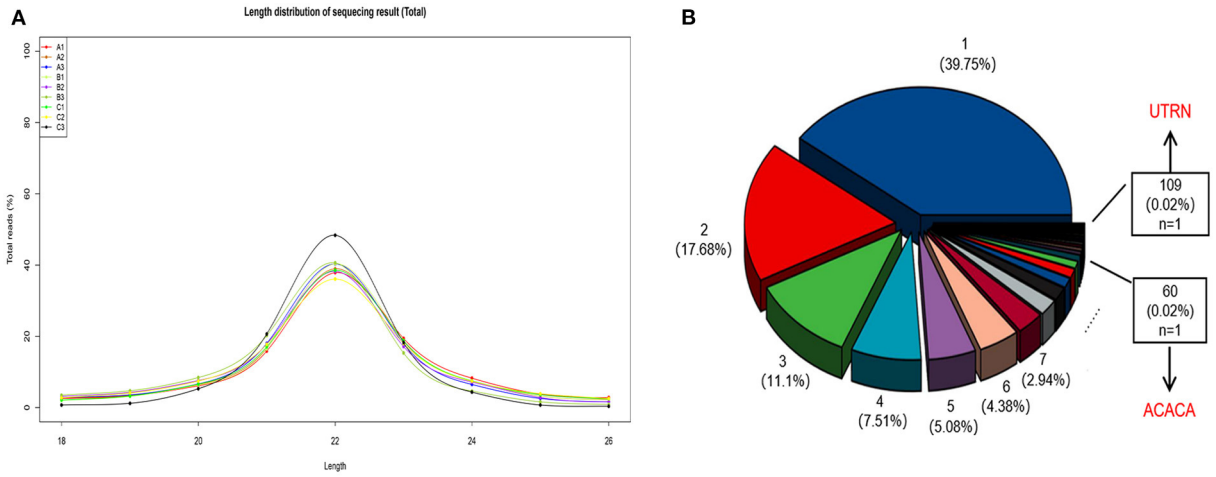


FIGURE 1
(A) The length distribution of sequenced miRNAs. 6M: A1, A2, A3; 18M: B1, B2, B3; 30M: C1, C2, C3. **(B)** Percentage of circRNA host genes. For example, the blue pie chart shows that 39.75% of the genes are transcribed to form 1 circRNA, meanwhile the red pie chart shows that 17.68% of the genes can be transcribed to make 2 circRNAs, and so on.

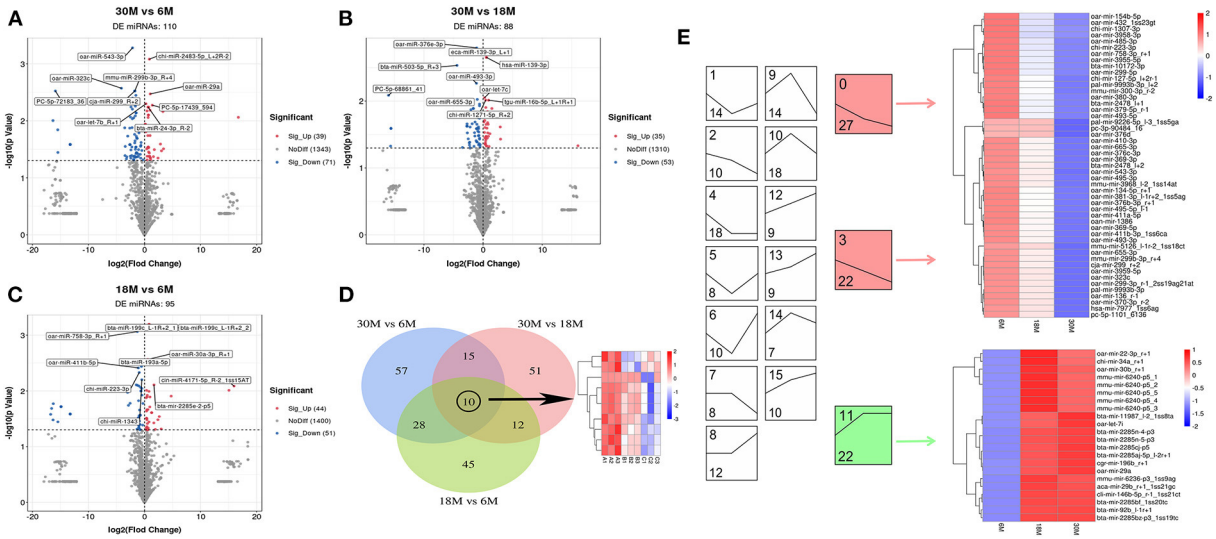


FIGURE 2
miRNA differential expression analysis. **(A–C)** The volcano plot of DE miRNA. **(A)** 30M vs. 6M; **(B)** 30M vs. 18M; **(C)** 18M vs. 6M. Annotated as the top five DE miRNA based on *P*-value. **(D)** Venn diagram analysis of DE miRNA. **(E)** Expression trend analysis of DE miRNA. The upper number indicates the ordinal number of each trend, while the lower number is the number of genes enriched, and those with color are the significantly enriched trends.

Differential expression analysis

We compared miRNAs (Supplementary Table S2) and circRNAs (Supplementary Table S3) in the tail fat of SS at three different stages (30 vs. 6 M, 30 vs. 18 M, and 18 vs. 6 M). The largest number of DE miRNAs were obtained for the 30 vs. 6 M comparison, with a total of 110 DE miRNAs (39 upregulated and 71 downregulated), including four novel DE miRNAs (Figures 2A–C). The 30 vs. 18 M comparison yielded 88 DE miRNAs (35 upregulated and 53 downregulated), including 3 novel DE miRNAs. For 18 vs. 6 M, 95 DE miRNAs (44 upregulated and 51 downregulated) were obtained, including seven novel DE miRNAs. Furthermore, 10 overlapping DE miRNAs were identified between the three comparison groups. These were highly expressed in 6 M SS, and their expression decreased with age (Figure 2D). Therefore, we postulated

that these DE miRNAs may be downregulated in parallel to the increase in tail fat, validating this hypothesis *via* trend enrichment analysis and heatmaps (Figure 2E). The DE miRNAs were enriched in 16 terms, three of which were significantly enriched ($P < 0.05$), colored in red and green. Forty-nine DE miRNAs were enriched with downregulated trends (red), and 22 DE miRNAs were enriched with upregulated trends (green), indicating that the DE miRNAs were mainly downregulated, which confirmed our hypothesis.

A total of 93 (62 upregulated and 31 downregulated), 89 (56 upregulated and 33 downregulated), and 66 (38 upregulated and 28 downregulated) DE circRNAs were obtained for 30 vs. 6 M, 30 vs. 18 M, and 18 vs. 6 M, respectively (Figures 3A–C). None of these were found to overlap among the three comparison groups (Figure 3D). Compared with DE miRNA, the expression trends in DE circRNAs were mainly enriched in the upregulation trend (green), indicating that these DE circRNAs may play a key role in the later stages of sheep fat tail growth (Figure 3E). Overall, the majority of DE miRNAs and circRNAs exhibited contrasting expression, indicating that the negative regulation between these molecules may play an important role in SS tail fat development.

Target gene prediction analysis

To build a ceRNA network of mRNA-miRNA-circRNA in sheep tail fat, miRNA target genes were predicted. To this end, we used TargetScan and miRanda, and all DE miRNAs (110) from the 30 vs. 6 M comparison were shown to bind to at least one mRNA. Four DE miRNAs were bound to more than 100 mRNAs. chi-miR-1343 targeted the most mRNAs (118), followed by oar-miR-370-3p_R-2 (111), bta-miR-2387_R+1 (102), and oan-miR-103-3p_R+2 (100). Among 30 vs. 18 M and 18 vs. 6 M, the novel PC-3p-43105_133 was the only DE miRNA targeting more than 100 mRNAs, 109 mRNAs, and 103 mRNAs, respectively. Thus, it could be a major regulator of fat tail development. Likewise, chi-miR-1343 was predicted to bind 88 mRNAs in 30 vs. 18 M and 83 mRNA in the 18 vs. 6 M comparison, suggestive of a multifaceted regulatory role.

In contrast, we predicted a binding interaction between DE miRNAs and DE circRNAs. For 30 vs. 6 M, there were 81 circRNAs bound to 99 miRNAs. Nine circRNAs were predicted to bind oar-miR-27a_R-1 and chi-miR-27b-3p. circRNA9695 was bound to the largest number of miRNAs (24), and three other circRNAs were bound to more than 10 miRNAs, namely, circRNA4175 (16), circRNA1985 (11), and circRNA382 (11). Likewise, these three circRNAs were found to bind multiple miRNAs from the 30 vs. 18 M comparison group. Thus, circRNA4175, circRNA1985, and circRNA382 could be major regulators of fat tail development. In 30 vs. 18 M, we found 71 circRNA and 71 miRNA targeting relationships, and the novel miRNA PC-3p-43105_133 was bound to the most circRNAs (13), once again highlighting the importance of this comparison group. Another miRNA family, miR-16a (chi-miR-16a-5p and oar-miR-16b_R+3), was found to bind to nine circRNAs. The 65 miRNAs were bound to 50 circRNAs in 18 vs. 6 M. mmu-miR-6240-p5 was found to bind the most mRNAs (9). Taken together, we predicted that multiple miRNAs bind to circRNAs that may play a key role in SS tail fat growth.

Validation of sequencing results

To validate RNA-seq results, we randomly selected six miRNAs and eight circRNAs and measured their expression in the 6, 18, and 30 M groups *via* qRT-PCR (Figure 4). The relative expression data determined *via* qRT-PCR were consistent with FPKM values obtained *via* RNA-seq, indicating that the RNA-seq data were reliable.

GO annotation and KEGG pathway analysis of miRNA target genes

The GO enrichment analysis of target genes was classified into biological process (BP), cellular component (CC), and molecular function (MF) categories (Supplementary Figure S1). More than half of the enriched GO terms for the three comparison groups were from the BP category, including the lowest number. CC terms were enriched with the highest number of genes in comparison groups, which mainly included the membrane, nucleus, and exosomes. Taken together, tail fat development is driven by diverse biological processes and cellular interactions. In three different stages, the miRNA target genes were significantly enriched for 843 GO terms, and the top 15 most significantly enriched GO terms are presented in a scatter plot (Figure 5A). Among the three comparison groups, we identified five GO terms that were enriched between two groups. These included actin cytoskeleton, intracellular membrane-bound organelles, extracellular space, extracellular exosome, and focal adhesion of the CC category, in addition to fatty acid ligase activity of the MF category. These miRNAs may play a significant regulatory role in fatty acid metabolism.

KEGG pathway analysis of the miRNA target genes was also performed, and the top 15 KEGG pathways are shown in a scatter plot (Figure 5B). A total of 19 KEGG pathways were significantly enriched. Among these, we noticed that the Rap1 signaling pathway, adherens junction, tight junction, cell adhesion molecules (CAMs), and regulation of actin cytoskeleton were enriched in the different growth stages. Furthermore, butanoate metabolism was enriched in both the 30 vs. 6 M and 30 vs. 18 M comparison groups.

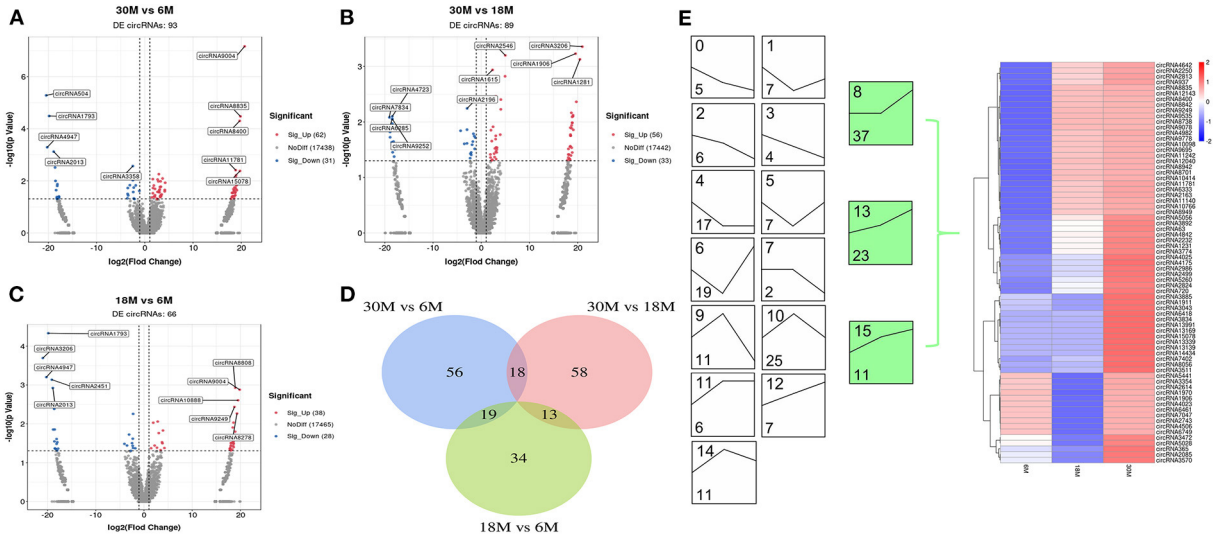


FIGURE 3
circRNA differential expression analysis. **(A–C)** Volcano plot of DE circRNAs. **(A)** 30M vs. 6M; **(B)** 30M vs. 18M; **(C)** 18M vs. 6M. Annotated as the top five DE circRNAs based on *P*-value. **(D)** Venn diagram analysis of DE circRNA. **(E)** DE circRNA expression trend analysis. The upper number indicates the ordinal number of each trend, while the lower number is the number of genes enriched, and those with color are the significantly enriched trends.

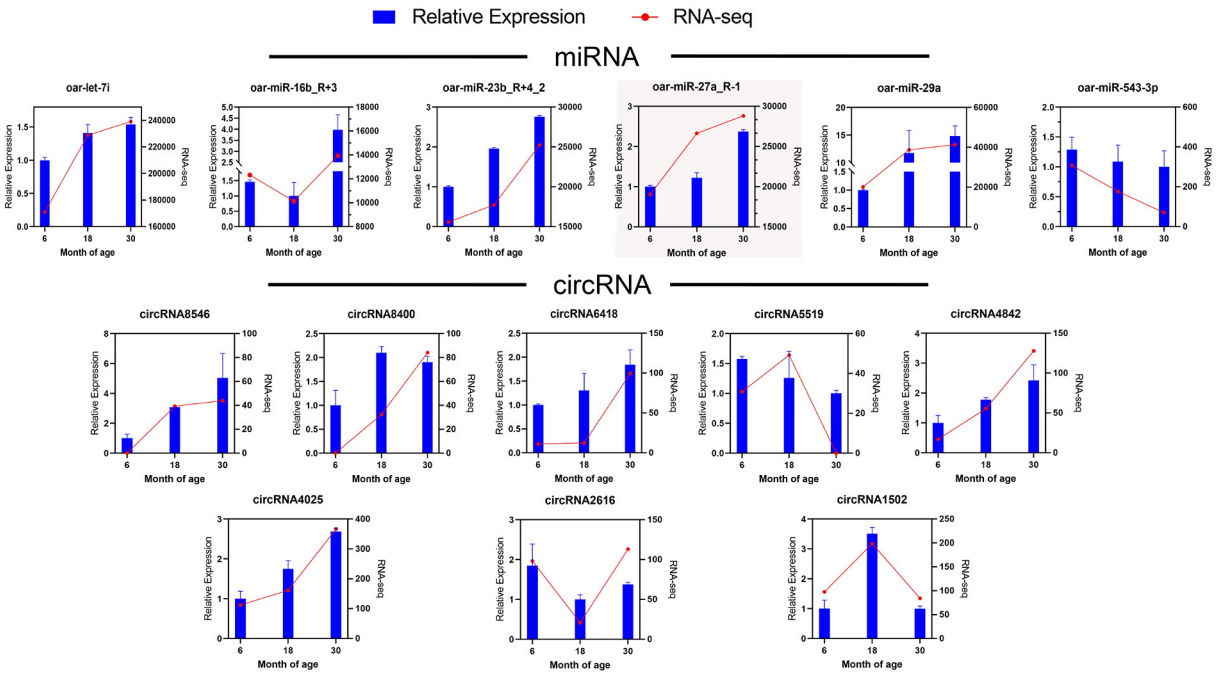


FIGURE 4
Validation of DE miRNAs and DE circRNAs *via* qRT-PCR. The blue bar represents qRT-PCR data, and the red line represents RNA-seq data.

Functional analysis of circRNA host genes

We employed GO and KEGG enrichment analysis to investigate the role of circRNAs in tail fat growth mediated *via* their host genes. Most enriched GO terms belonged to the BP category, and most of the genes were enriched in CC (Supplementary Figure S2). We obtained 350 significantly enriched GO terms in the three comparison groups. The top 15

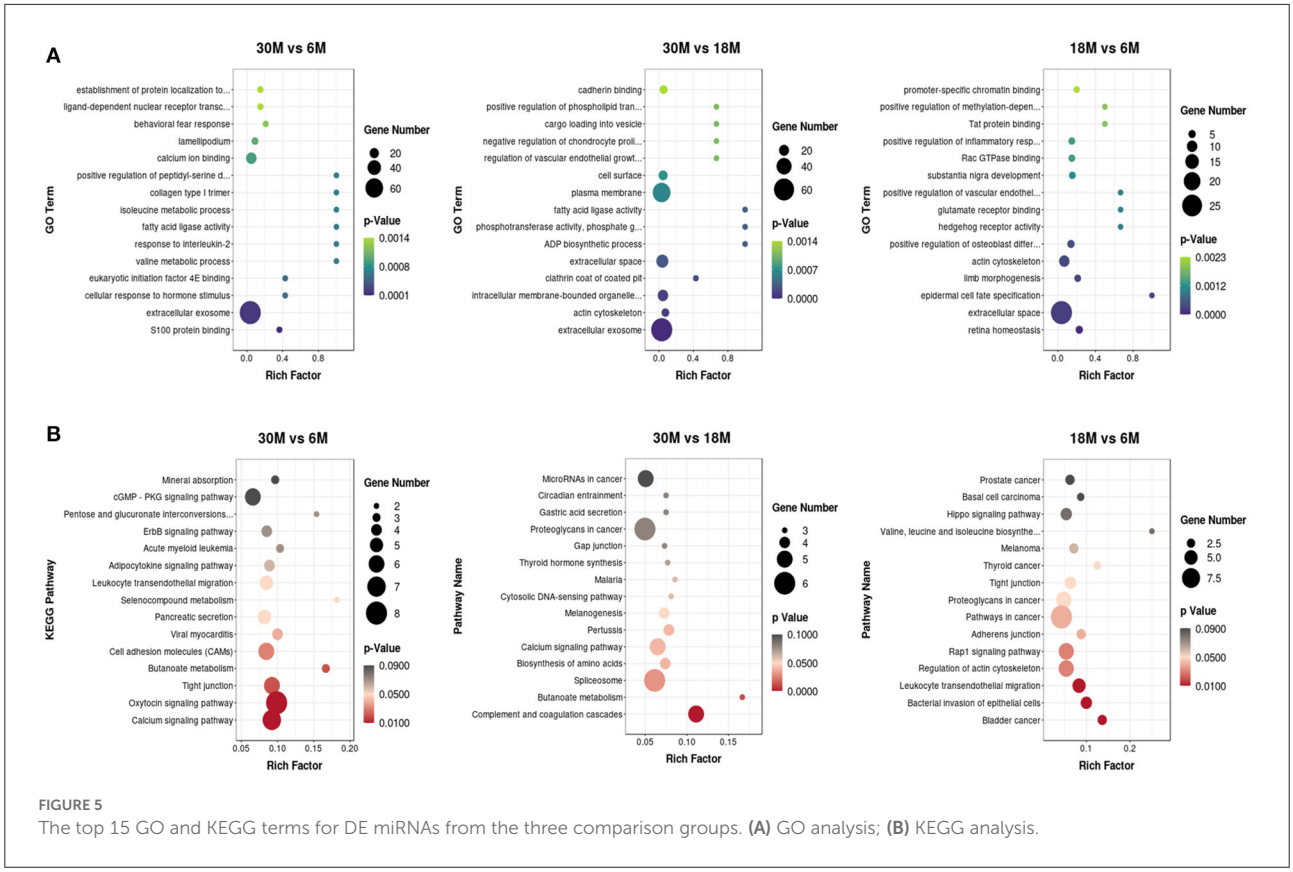


FIGURE 5
The top 15 GO and KEGG terms for DE miRNAs from the three comparison groups. **(A)** GO analysis; **(B)** KEGG analysis.

GO terms are presented in a scatter plot (Figure 6A). GTPase activator activity was enriched in both the 30 vs. 6 M and 30 vs. 18 M comparison groups, and fatty acid beta-oxidation was among the significantly enriched GO terms for 18 vs. 6 M. The top 15 enriched KEGG pathways are shown in Figure 6B. There were 22 significantly enriched KEGG pathways, among which several fat-related pathways were identified. These processes included propanoate metabolism, fatty acid metabolism, fatty acid biosynthesis, unsaturated fatty acid biosynthesis, and fatty acid elongation. Fatty acid metabolism was the only pathway enriched in all three groups. Host genes ACACA (circRNA382, circRNA392, and circRNA394) and HADHA (circRNA10888) were enriched in these pathways, indicative of the involvement of these circRNAs in fatty acid metabolism within tail fat.

Construction of the mRNA-miRNA-circRNA co-expression network

Using Cytoscape (version 3.9.0), we constructed co-expression networks based on the mRNAs, miRNAs, and circRNAs identified in sheep fat tails (Figure 7). In this manner, we determined co-expression relationships based on the negative correlation between miRNA expression and target gene expression, given the targeting interactions between miRNA (sheep species), mRNA, and circRNA. A total of 35 (down-up-down) and 19 (up-down-up) mRNA-miRNA-circRNA co-expression patterns were obtained. This indicates that the down-up-down co-expression predominated the network, suggesting that upregulated miRNAs in the ceRNA network play a central regulatory role in tail fat. It is worth noting that oar-miR-16b_R+3 was targeted to two mRNAs and seven circRNAs, indicating that oar-miR-16b_R+3 also has regulatory functions with varied expression patterns at different stages of tail fat development. More importantly, we discovered that oar-miR-27a_R-1 and oar-miR-29a are at the center of the regulatory network and bind to several mRNAs and circRNAs, including TKT, ACSL4, GPAM, and POSTN. Among them, GPAM and ACSL4 are closely related to fat metabolism. Therefore, we suggest that oar-miR-27a_R-1 and oar-miR-29a have an important significance in the overall ceRNA relationship, and play a key role in the growth and metabolism of sheep tail fat.

miRNA target validation

In our study, the dual-luciferase reporter system was used to verify the relationship between miRNAs and their targets.

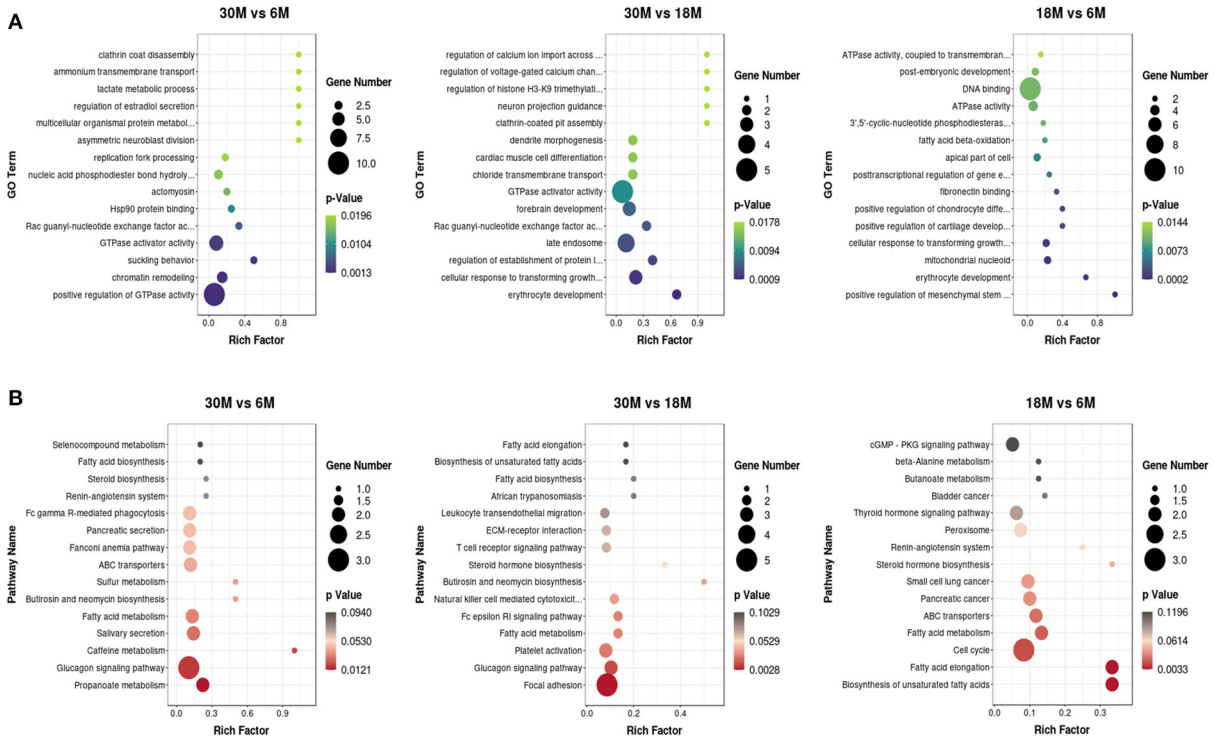


FIGURE 6
The top 15 enriched GO and KEGG terms for DE circRNAs. **(A)** GO analysis; **(B)** KEGG analysis.

We further analyzed sheep DE miRNAs and their targets to identify ceRNAs that may be associated with tail fat growth. We filtered highly expressed sheep miRNAs (average norm value > 1,000) and sought to identify miRNAs that had a major effect on sheep tail fat development (Supplementary Table S4). Among these highly expressed miRNAs, oar-miR-27a_R-1 and oar-miR-29a were included, and the analysis above revealed that these genes were critical for the growth and metabolism of sheep tail fat. This led us to select these two miRNAs as the validation miRNAs. According to the ceRNA co-expression network, oar-miR-27a_R-1 was upregulated at 30 vs. 6 M and targeted to ACSL4, which is related to fatty acid metabolism. circRNA1985 was downregulated at 30 vs. 6 M, binding to oar-miR-27a_R-1. oar-miR-29a was upregulated at 30 vs. 6 M and 18 vs. 6 M, binding with 5 DEGs. Among them, GPAM was related to glycerolipid and glycerophospholipid metabolism. As a competitive binding RNA, circRNA3539 was downregulated in this process. Therefore, we sought to validate the ceRNA regulatory relationship between oar-miR-27a_R-1 and oar-miR-29a as the core, that is, between ACSL-oar-miR-27a_R-1-circRNA1985 and GPAM-oar-miR-29a-circRNA3539.

We found only one target site in all the targeted pairs (Figure 8). Furthermore, compared with the negative control (NC) group, oar-miR-27a_R-1 and oar-miR-29a significantly downregulated the expression of luciferase in their predicted WT (wild type) targets ($P < 0.001$), indicating that there is a binding effect between all predicted targets in our study. After mutation, these three miRNAs failed to downregulate the expression of luciferase of mutant targets ($P > 0.05$) compared with the NC group, indicating that mutation was successful. Taken together, these results suggest that oar-miR-27a_R-1 can decrease ACSL4 expression by targeting the ACSL4-3′-UTR, and circRNA1985 can competitively bind with oar-miR-27a_R-1 and thus regulate the expression of ACSL4. The same conclusion was drawn for GPAM-oar-miR-29a-circRNA3539.

Discussion

As a part of the sheep tissue, the tail fat can store heat for the sheep's body and help them resist the cold winter. As a by-product of mutton, tail fat provides the energy needed by the human body, and it can also be used as a raw material for human daily necessities, such as soap and medicinal materials. At present, sheep tail fat is gradually entering the field of vision of researchers. Previous studies have conducted transcriptome analysis on different breeds of sheep tails, and the results may be affected by the breed effect to some extent (26). However, the amount of intramuscular fat has been shown in previous study

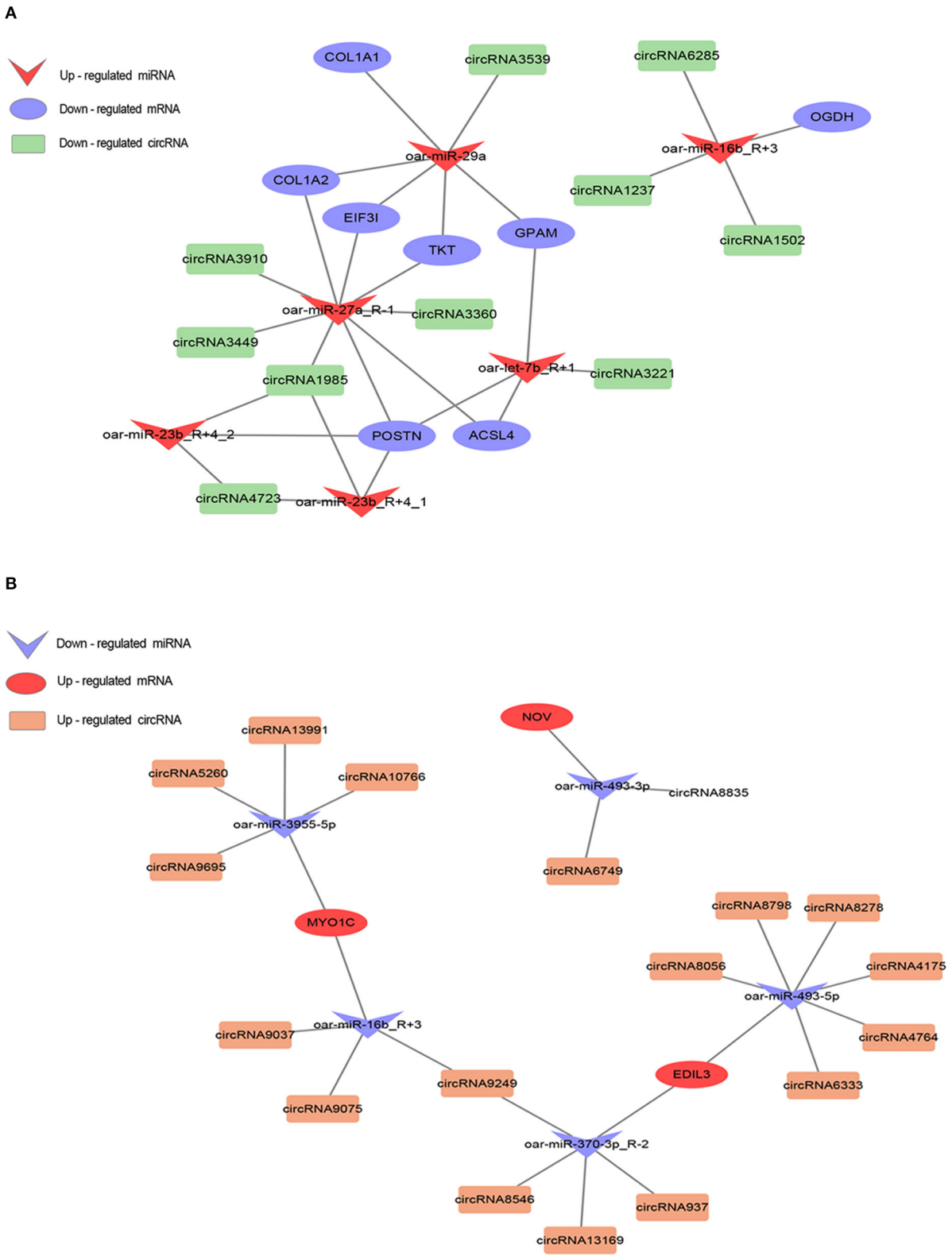


FIGURE 7
ceRNA regulatory network analysis in sheep tail fat. **(A)** Down-up-down mode. **(B)** Up-down-up mode. The shapes represent different RNAs and the colors represent different regulations.

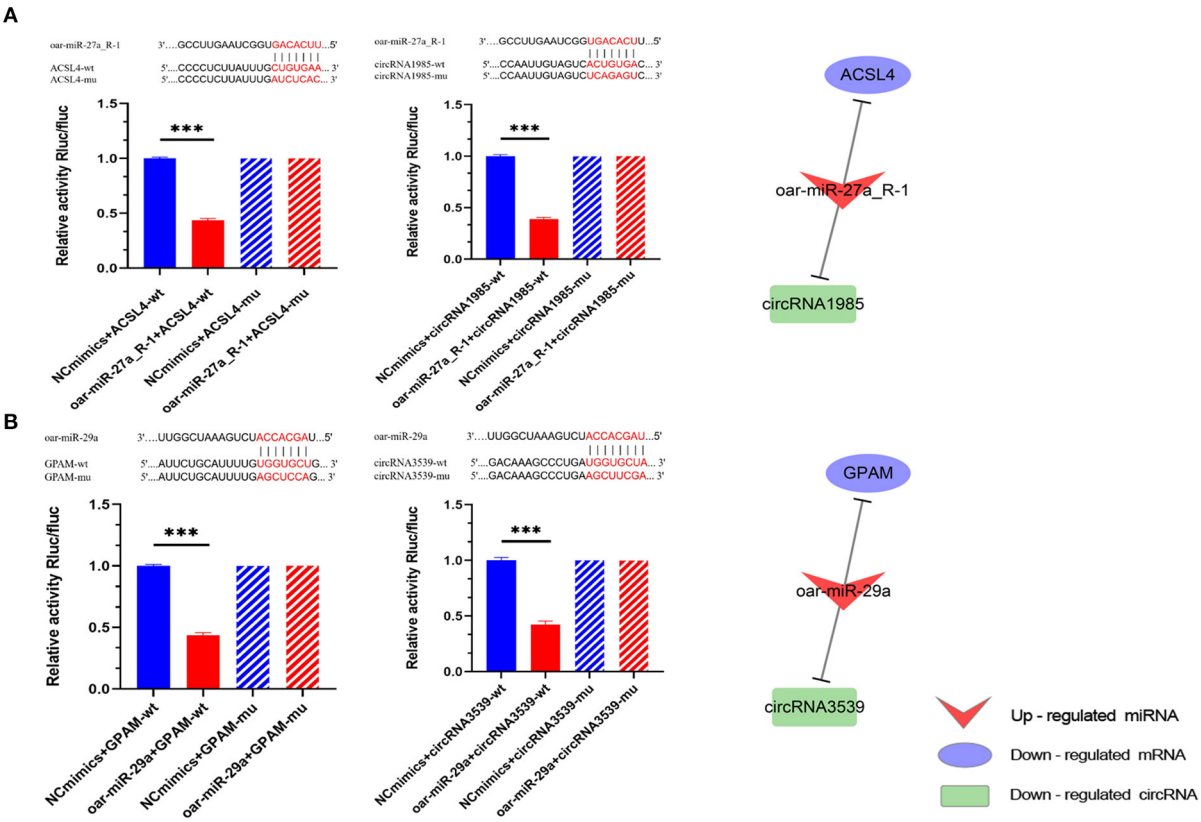


FIGURE 8
ceRNA dual-luciferase reporter gene analysis. **(A)** Binding site validation for oar-miR-27a_R-1 and ACSL4 as well as oar-27a_R-1 for circRNA1985; **(B)** Binding site validation for oar-miR-29a and GPAM as well as oar-miR-29a and circRNA3539. Bars without slashes show the result of inserting the mutant sequence into the plasmid, while solid bars show the result of inserting the original sequence. The group with extra miRNA sequences is shown in red, and the group without any miRNA sequences (control group) is shown in blue. ***$P < 0.001$.

to be significantly impacted by aging (28). Generally, Sunite fat-tailed sheep mainly start slaughter at the age of 6 months, and after growth and development throughout the year, the tail fat increases continuously and reaches the weight of about 3–4.5 kg at 30 months of age. Therefore, in our study, we used the fat tail from Sunite sheep in three different stages, namely 6, 18, and 30 months to understand the potential molecular mechanism in fat development. The miRNAs and circRNAs were obtained by RNA-seq from the fat tail tissues of the three groups. Then, we explored the potential mechanism of sheep tail fat regulation in three different stages through functional analysis, and finally highlighted ceRNA regulation by the construction of co-expression networks and miRNA target validation.

With the advancement of RNA deep sequencing technology, miRNA-seq technology is increasingly being used in various animal species, resulting in the discovery of a large number of novel miRNAs. In this study, miRNAs from closely similar species found in miRBase were compared and identified during the analysis in order to more comprehensively refer to the registration data of miRBase. In spite of this, out of 1,942 miRNAs, we still found 392 novel miRNAs. The identification of these novel miRNAs may inspire new avenues of research in related areas, and additional research may provide insight on their biological significance in regulating lipid metabolism in sheep tail adipose tissue.

A total of 219 DE miRNAs (including 12 novel miRNAs) were detected in three comparison groups of different growth stages in sheep fat tail. Among these DE miRNAs, each comparison group had DE miRNAs that are uniquely expressed, and there were DE miRNAs that are commonly expressed in any two comparison groups or in three groups. We found that these DE miRNAs were mainly down-regulated and that their expression decreased with sheep age, which was supported by our expression trend enrichment analysis. Previous studies in Han sheep adipose tissue yielded similar results, and the mechanism underlying their down-regulation remains unknown (29).

CeRNAs regulate mRNA expression *via* competitively binding to miRNAs (30). We used TargetScan and miRanda to predict miRNAs target genes, and results showed that

110 DE miRNAs were predicted to bind to at least one mRNA. Among them, 5 DE miRNA were bound to more than 100 mRNAs, namely oar-miR-370-3p_R-2, bta-miR-2387_R+1, oan-miR-103-3p_R+2, chi-miR-1343, and PC-3p-43105_133. A previous study indicated that miRNA-1343-5p was predicted to bind to the key adipogenic gene C/EBP in bovine adipocytes (31). *In vivo* and *in vitro* studies suggest that miR-370 may alter fatty acid composition during adipogenesis, promote 3T3-L1 preadipocyte proliferation, and inhibit differentiation by directly targeting Mknk1 (32). Further, miR-370 may promote the expression of lipogenic genes SREBP-1c, DGAT2, FAS, and ACC1 (33). miR-103 has been described as playing a critical role in lipid metabolism. Studies have shown that miR-103 inhibits the expression of FASN and SCD1 *via* direct binding, in addition to promoting the differentiation of 3T3-L1 cells by targeting MEF2D and activating the AKT/mTOR signaling pathway (34, 35). miR-1343 may play a key role in sheep intestinal tissue based on PIK3R1 being its predicted target gene (36). In addition, chi-miR-1343 was also found to bind to multiple mRNAs in the 30 vs. 18 M and 18 vs. 6 M groups. Thus, we suggest that chi-miR-1343 may play an important regulatory role in sheep tail fat growth and development, and its mechanism in metabolic regulation needs to be further elucidated. Further, the novel PC-3p-43105_133 was the only DE miRNA targeting more than 100 mRNAs among the 30 vs. 18 M and 18 vs. 6 M groups. The novel miRNAs might represent a crucial regulatory component. For instance, recent research has demonstrated that the novel miRNA Y-56 targets IGF-1R to regulate the proliferation and cell cycle processes of porcine skeletal muscle satellite cells (37). However, there is still a limitation of novel miRNAs related to adipose tissue, and the mechanisms of many novel miRNAs remain unknown. As a result, it is essential to investigate the mechanisms of these novel miRNAs.

To uncover the functions of these DE miRNAs in tail adipose tissue, we utilized GO enrichment and KEGG pathway analyses. Among the three comparison groups, we found that GO functions of the actin cytoskeleton, intracellular membrane-bounded organelles, extracellular space, extracellular exosome, and focal adhesion of the CC category, and fatty acid ligase activity of the MF category, were enriched between two groups. Studies have shown that the inhibition of focal adhesion kinase (FAK) leads to an upregulation of adipogenic marker genes AP2 and LEP and lipid accumulation (38). The FATP family of proteins, which have fatty acid acyl-CoA ligase activity, is present in the plasma membrane and intracellular organelles (39). However, none of the miRNAs were found to bind to any of the genes in the FATP family in our study. Instead, ACSM1 and ACSM3, which are related to fatty acid ligase activity (http://geneontology.org/), were found to bind 24 miRNAs. These miRNAs may play a significant regulatory role in fatty acid metabolism. We also found several significant pathways in the KEGG pathway analysis, including the Rap1 signaling pathway, adherens junction, tight junction, cell adhesion molecules (CAMs), regulation of actin cytoskeleton, and butanoate metabolism. Tight junctions and adherens junctions are two types of cellular junctions that fundamentally affect cell proliferation and differentiation (40). Actin is a major component of adherens junctions, and abnormalities in this cytoskeletal protein impede the assembly of adherens junctions (41). Rap1 is mainly implicated in the control of cell adhesion, cell junction formation, secretion, and cell polarity (42). The significant enrichment of these pathways suggests that DE miRNAs may play prominent roles in sheep tail adipogenesis through cell-to-cell interactions. In addition, no studies have suggested a role for the Rap1 signaling pathway in adipose tissue. The present study may provide a theoretical basis for future research in this direction. Butyrate is processed by acetyl-CoA to generate fatty acids, cholesterol, and ketone bodies, thus providing specialized substrates for lipid biosynthesis (43). Butyrate treatment can cause adipocytes to accumulate triglycerides *in vitro* (44). This indicates that the miRNA-mediated regulation of butyric acid may be a potential regulator of SS tail adipose tissue growth.

MiRNAs have several classical molecular regulatory mechanisms, and in addition to binding to target mRNA transcripts *via* complementary base pairing, they can also target circRNAs to exert negative regulatory effects. Out of the 17,531 identified circRNAs, 198 DE circRNAs were screened in the current study. We found that these circRNAs were mainly up-regulated using expression trend enrichment analysis, which was the opposite of the main trend of miRNAs, which appeared to be negatively regulated. Therefore, we suggest that negative regulation between these molecules may be essential for the production of SS tail fat. To elucidate the regulatory relationship between miRNAs and circRNAs, we predicted the linkage between them. In the 30 vs. 6 M and 30 vs. 18 M comparative groups, it was observed that circRNA4175, circRNA1985, and circRNA382 bind to numerous miRNAs, which is noteworthy. In addition, we discovered that, among the three comparison groups, three miRNA families bind to the multiple circRNAs. They are miR-16 (chi-miR-16a-5p and oar-miR-16b_R+3) at 30 vs. 18 M, mmu-miR-6240-p5 at 18 vs. 6 M, and miR-27 (oar-miR-27a_R-1 and chi-miR-27b-3p) at 30 vs. 6 M. miR-27 is among the many miRNAs involved in cholesterol homeostasis and fatty acid metabolism. miR-27a and−27b were found to suppress adipocyte differentiation by regulating peroxisome proliferator-activated receptor γ (45). Previous research found that overexpression of miR-16a-5p can promote the expression of adipogenic marker genes as well as 3T3-L1 adipocyte differentiation by binding to the EPT1 gene (46). Current research on miR-6240 has focused on human heart disease, with one study suggesting that the downregulation of miR-6240 leads to an increase in white adipocyte markers (47). We also found that three

circRNAs could bind to any two of these three miRNAs families, but only miR-27 could bind to all three. We therefore propose that the miR-27 family occupies a more central role in the interaction with circRNAs in the sheep tail adipose tissue.

In addition to regulating gene expression through competitive binding with miRNAs, circRNAs can also regulate the expression of their host genes (20). In the GO functional analysis of the circRNA host genes, we discovered that GTPase activator activity was enriched in both the 30 vs. 6 M and 30 vs. 18 M comparison groups. Recent studies have shown that GTPases play an important regulatory role in adipogenic differentiation (48, 49). Fatty acid beta-oxidation, which is crucial for the maintenance of thermogenesis, was among the significantly enriched GO terms for 18 vs. 6 M (50). KEGG pathway analysis suggested multiple adipose-related pathways were emphasized in our study. Among them, fatty acid metabolism was the only pathway enriched in all three groups, and host gene ACACA was enriched in these pathways. Combined with the findings of the circRNA host gene analysis (60 circRNAs from ACACA, Figure 1), we suggest that the majority of the circRNAs play a key metabolic role in tail fat.

Finally, we highlighted the potential regulatory mechanisms of ceRNAs in SS tail fat metabolism based on previous analysis of the expression profiles of miRNAs and circRNAs. In our study, we constructed two co-expression networks based on the mRNAs, miRNAs, and circRNAs identified in sheep fat tails. A total of 35 (down-up-down) and 19 (up-down-up) mRNA-miRNA-circRNA co-expression patterns were obtained. Herein, oar-miR-27a_R-1 was linked to multiple target genes, including TKT, ACSL4, and POSTN, which are involved in fatty acid metabolism and adipogenesis (50–52). As previously stated, miR-27 is a major regulator within adipose tissue, and miR-27a has been associated with lipid accumulation differences between intramuscular and subcutaneous adipose tissue in sheep (53). In the present study, miR-27a targeted multiple mRNAs and was shown to be regulated by up to four circRNAs. Therefore, we suggest that this miRNA may be a critical regulator of sheep tail fat development. Furthermore, oar-miR-29a binds to multiple mRNAs and circRNAs in the co-expression network. miR-29a was reported to associate with multiple biological processes, including lipid metabolism (54). In addition to TKT, miR-29a also targets to GPAM. GPAM was related to glycerolipid and glycerophospholipid metabolism, and its overexpression leads to the increase of triglyceride levels and lipid metabolism-related gene expression (55). Combined with the expression level of these two miRNAs, we suggest that oar-miR-29a and oar-miR-27a_R-1 play an important role in the ceRNA network. Therefore, we constructed the ceRNA interactions ACSL-oar-miR-27a_R-1-circRNA1985 and GPAM-oar-miR-29a-circRNA3539 using these two miRNAs as the core, and we successfully validated their interactions using dual luciferase gene reporter analyses. As a result, we verified two ceRNA expression networks and suggested multiple functions related to fat metabolism in sheep tail fat development. However, the involvement of these ceRNAs in sheep tail-fat metabolism requires further investigation. Overall, the investigation into the expression profiles of miRNAs and circRNAs at different developmental stages fills a research gaps in the study of the Sunite sheep tail fat metabolic mechanisms and provides new thoughts for future studies. For instance, these studies could focus on exploring the regulatory mechanisms of the novel miRNAs, the circRNAs that regulate their host genes, and the negative regulation of miRNAs, and on investigating more potential regulatory mechanisms for ceRNAs. The current results provide a theoretical basis for the identification of molecular markers related to sheep tail fat metabolism.

Conclusions

In this study, we established miRNA and circRNA expression profiles to investigate the potential regulatory mechanisms underlying tail fat development in SS. At different growth stages, DE miRNAs may play a role in cell-to-cell interactions through target binding. The host genes of DE circRNAs were shown to be more involved in lipid and fatty acid metabolism. Based on the target prediction study, we created a miRNA-centered ceRNA regulatory network and filtered critical miRNAs for the validation of multiple target loci. Among them, we highlight the relationship pair of ACSL-oar-miR-27a_R-1-circRNA1985 and GPAM-oar-miR-29a-circRNA3539 ceRNAs centered on oar-miR-27a_R-1 and oar-miR-29a, revealing potential ceRNA networks involved in the regulation of tail fat development. Our findings highlight potential ceRNAs involved in sheep tail fat development and provide a theoretical basis for by-product utilization.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

All experimental procedures were approved by the Animal Ethics Committee of the Inner Mongolia Agricultural University Animal Experimentation Area and followed the Chinese Animal Protection Law. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

XH and GB designed the study. XH, LC, YHa, and YHu performed the experiment. XH, RW, YY, and XQ analyzed the data. JW, LS, and GB provided ideas and suggestions. XH wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the China Agriculture Research System of the MOF and the MARA (Grant No. CARS38) as well as the Inner Mongolia Autonomous Region Science and Technology Plan Project (2019–2022).

Acknowledgments

We appreciate the generous help received from the Laboratory of Meat Science and Biotechnology of the Inner Mongolia Agricultural University, China.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.954882/full#supplementary-material

SUPPLEMENTARY FIGURE S1
GO function classification of DE miRNAs. **(A)** 30 vs. 6 M; **(B)** 30 vs. 18 M; **(C)** 18 vs. 6 M.

SUPPLEMENTARY FIGURE S2
GO function classification of DE circRNAs host genes. **(A)** 30 vs. 6 M; **(B)** 30 vs. 18 M; **(C)** 18 vs. 6 M.

SUPPLEMENTARY TABLE S1
qRT-PCR primer sequences.

SUPPLEMENTARY TABLE S2
The differentially expressed miRNA.

SUPPLEMENTARY TABLE S3
The differentially expressed circRNA.

SUPPLEMENTARY TABLE S4
The highly expressed DE miRNA in tail fat.

References

1. Kashan NE J, Azar GHM, Afzalzadeh A, Salehi A. Growth performance and carcass quality of fattening lambs from fat-tailed and tailed sheep breeds. *Small Ruminant Res.* (2005) 60:267–71. doi: 10.1016/j.smallrumres.2005.01.001

2. Pan YY, Jing JJ, Zhao JX, Jia XL, Qiao LY, An LX, et al. MicroRNA expression patterns in tail fat of different breeds of sheep. *Livest Sci.* (2018) 207:7–14. doi: 10.1016/j.livsci.2017.11.007

3. Zhou GX, Wang XL, Yuan C, Kang DJ, Xu XC, Zhou JP, et al. Integrating miRNA and mRNA expression profiling uncovers miRNAs underlying fat deposition in sheep. *Biomed Res Int.* (2017) 2017:1857580. doi: 10.1155/2017/1857580

4. Han FH, Zhou LS, Zhao L, Wang L, Liu LR, Li HJ, et al. Identification of miRNA in sheep intramuscular fat and the role of miR-193a-5p in proliferation and differentiation of 3T3-L1. *Front Genet.* (2021) 12:633295. doi: 10.3389/fgene.2021.633295

5. Li TT, Wang X, Luo RR, An XJ, Zhang Y, Zhao XX, et al. Integrating miRNA and mRNA profiling to assess the potential miRNA-mRNA modules linked with testicular immune homeostasis in sheep. *Front Vet Sci.* (2021) 8:647153. doi: 10.3389/fvets.2021.647153

6. Yang H, Fu L, Luo QF, Li LC, Zheng FL, Wen JY, et al. Identification and validation of key miRNAs and miRNA-mRNA regulatory network associated with uterine involution in postpartum Kazakh sheep. *Arch Anim Breed.* (2021) 64:119–29. doi: 10.5194/aab-64-119-2021

7. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature.* (2013) 495:384–8. doi: 10.1038/nature11993

8. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE.* (2012) 7:e30733. doi: 10.1371/journal.pone.0030733

9. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu JZ, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA.* (2013) 19:141–57. doi: 10.1261/rna.035667.112

10. Li A, Huang W, Zhang X, Xie L, Miao X. Identification and characterization of circRNAs of two pig breeds as a new biomarker in metabolism-related diseases. *Cell Physiol Biochem.* (2018) 47:2458–70. doi: 10.1159/000491619

11. Huang JP, Zhao JH, Zheng QZ, Wang SZ, Wei XF, Li F, et al. Characterization of circular RNAs in Chinese buffalo (Bubalusbubalis) adipose tissue: a focus on circular RNAs involved in fat deposition. *Animals.* (2019) 9:403. doi: 10.3390/ani9070403

12. He XG, Wu RH, Yun YY, Qin X, Chen L, Han YF, et al. Transcriptome analysis of messenger RNA and long noncoding RNA related to different developmental stages of tail adipose tissues of sunite sheep. *Food Sci Nutr.* (2021) 9:5722–34. doi: 10.1002/fsn3.2537

13. Sassi-Zaidy YB, Maretto F, Zanetti E, Hajji GM, Cassandro M. Genetic structure and variability within and among populations of the fat-tailed barbarine sheep breed using microsatellites markers. *Afr J Biotechnol.* (2014) 13:44–54. doi: 10.5897/AJB2013.13363

14. Zhang W, Xu MS, Wang JJ, Wang SY, Wang XH, Yang JQ, et al. Comparative transcriptome analysis of key genes and pathways activated in response to fat deposition in two sheep breeds with distinct tail phenotype. *Front Genet.* (2021) 12:639030. doi: 10.3389/fgene.2021.639030

15. Wei CH, Wang HH, Liu G, Wu MM, Cao JX, Liu Z, et al. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics.* (2015) 16:194. doi: 10.1186/s12864-015-1384-9

16. Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* (2012) 13:10. doi: 10.1186/1471-2156-13-10

17. Wang SH, Liu SR, Yuan TT, Sun XZ. Genetic effects of FTO gene insertion/deletion (InDel) on fat-tail measurements and growth traits in Tong sheep. *Anim Biotechnol.* (2021) 32:229–39. doi: 10.1080/10495398.2019.1680379

18. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* (2016) 26:1277–87. doi: 10.1101/gr.202895.115

19. Zhang XO, Wang HB, Zhang Y, Lu XH, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell.* (2014) 159:134–47. doi: 10.1016/j.cell.2014.09.001

20. Shang FZ, Wang Y, Ma R, Di ZY, Wu ZH, Hai EH, et al. Expression profiling and functional analysis of circular RNAs in inner mongolian cashmere goat hair follicles. *Front Genet.* (2021) 12:678825. doi: 10.3389/fgene.2021.678825

21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

22. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics.* (2006) 7:191. doi: 10.1186/1471-2105-7-191

23. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* (2015) 4:e05005. doi: 10.7554/eLife.05005

24. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* (2010) 11:R90. doi: 10.1186/gb-2010-11-8-r90

25. Bakhtiarizadeh MR, Salami SA. Identification and expression analysis of long noncoding RNAs in fat-tail of sheep breeds. *G3.* (2019) 9:1263–76. doi: 10.1534/g3.118.201014

26. Ma L, Zhang M, Jin YY, Erdenee S, Hu LY, Chen H, et al. Comparative transcriptome profiling of mRNA and lncRNA related to tail adipose tissues of sheep. *Front Genet.* (2018) 9:365. doi: 10.3389/fgene.2018.00365

27. Ding RR, Yang M, Quan JP, Li SY, Zhuang ZW, Zhou SP, et al. Single-locus and multi-locus genome-wide association studies for intramuscular fat in duroc pigs. *Front Genet.* (2019) 10:619. doi: 10.3389/fgene.2019.00619

28. Bao G, Liu X, Wang J, Hu J, Shi B, Li S, et al. Effects of slaughter age on myosin heavy chain isoforms, muscle fibers, fatty acids, and meat quality in longissimus thoracis muscle of tibetan sheep. *Front Vet Sci.* (2021) 8:689589. doi: 10.3389/fvets.2021.689589

29. Miao X, Luo Q, Qin X, Guo Y. Genome-wide analysis of microRNAs identifies the lipid metabolism pathway to be a defining factor in adipose tissue from different sheep. *Sci Rep.* (2016) 5:18470. doi: 10.1038/srep18470

30. Zhao RR Li J, Liu N, Li HG, Liu LR, Yang F, et al. Transcriptomic analysis reveals the involvement of lncRNA-miRNA-mRNA networks in hair follicle induction in aohan fine wool sheep skin. *Front Genet.* (2020) 11:590. doi: 10.3389/fgene.2020.00590

31. Yu X, Fang XB, Gao M, Mi JQ, Zhang XQ, Xia LX, et al. Isolation and identification of bovine preadipocytes and screening of microRNAs associated with adipogenesis. *Animals.* (2020) 10:818. doi: 10.3390/ani10050818

32. Zhang PW, Li XR, Zhang SH, Wu S, Xiao Q, Gu Y, et al. miR-370-3p regulates adipogenesis through targeting Mknk1. *Molecules.* (2021) 26:6926. doi: 10.3390/molecules26226926

33. Qu Q, Zeng F, Liu X, Wang QJ, Deng F. Fatty acid oxidation and carnitine palmitoyltransferase I: emerging therapeutic targets in cancer. *Cell Death Dis.* (2016) 7:e2226. doi: 10.1038/cddis.2016.132

34. Li MH, Liu ZJ, Zhang ZZ, Liu GN, Sun SD, Sun C. miR-103 promotes 3T3-L1 cell adipogenesis through AKT/mTOR signal pathway with its target being MEF2D. *Biol Chem.* (2015) 396:235–44. doi: 10.1515/hsz-2014-0241

35. Zhang MY, Tang Y, Tang E, Lu WR. MicroRNA-103 represses hepatic *de novo* lipogenesis and alleviates NAFLD *via* targeting FASN and SCD1. *Biochem Biophys Res Commun.* (2020) 524:716–22. doi: 10.1016/j.bbrc.2020.01.143

36. Hou L, Ji ZB, Wang GZ, Wang J, Chao TL, Wang JM. Identification and characterization of microRNAs in the intestinal tissues of sheep (Ovis aries). *PLoS ONE.* (2018) 13:e0193371. doi: 10.1371/journal.pone.0193371

37. Song J, Hao L, Zeng X, Yang R, Qiao S, Wang C, et al. A Novel miRNA Y-56 targeting IGF-1R mediates the proliferation of porcine skeletal muscle satellite cells through AKT and ERK pathways. *Front Vet Sci.* (2022) 9:754435. doi: 10.3389/fvets.2022.754435

38. Hyväri L, Ojansivu M, Juntunen M, Kartasalo K, Miettinen S, Vanhatupa S. Focal adhesion kinase and ROCK signaling are switch-like regulators of human adipose stem cell differentiation towards osteogenic and adipogenic lineages. *Stem Cells Int.* (2018) 2018:2190657. doi: 10.1155/2018/2190657

39. Adeshakin AO, Liu W, Adeshakin FO, Afolabi LO, Zhang MQ, Zhang GZ, et al. Regulation of ROS in myeloid-derived suppressor cells through targeting fatty acid transport protein 2 enhanced anti-PD-L1 tumor immunotherapy. *Cell Immunol.* (2021) 362:104286. doi: 10.1016/j.cellimm.2021.104286

40. Pollard TD, Earnshaw WC, Lippincott-Schwartz J, Johnson GT. Cellular adhesion and the extracellular matrix, 31-intercellular junctions. In: *Cell Biology.* 3rd ed. Elsevier (2017). p. 543–53.

41. Dowland SN, Madawala RJ, Lindsay LA, Murphy CR. The adherens junction is lost during normal pregnancy but not during ovarian hyperstimulated pregnancy. *Acta Histochem.* (2016) 118:137–43. doi: 10.1016/j.acthis.2015.12.004

42. Raaijmakers JH, Bos JL. Specificity in Ras and Rap signaling. *J Biol Chem.* (2009) 284:10995–9. doi: 10.1074/jbc.R800061200

43. Liu H, Wang J, He T, Becker S, Zhang GL, Li DF, et al. Butyrate: a double-edged sword for health? *Adv Nutr.* (2018) 9:21–9. doi: 10.1093/advances/nmx009

44. Yan H, Ajuwon KM. Mechanism of butyrate stimulation of triglyceride storage and adipokine expression during adipogenic differentiation of porcine stromovascular cells. *PLoS ONE.* (2015) 10:e0145940. doi: 10.1371/journal.pone.0145940

45. Kim SY, Kim AY, Lee HW, Son YH, Lee GY, Lee JW, et al. miR-27a is a negative regulator of adipocyte differentiation *via* suppressing PPARγ expression. *Biochem Biophys Res Commun.* (2010) 392:323–8. doi: 10.1016/j.bbrc.2010.01.012

46. Xu JJ, Zhang LP, Shu GB, Wang B. microRNA-16–5p promotes 3T3-L1 adipocyte differentiation through regulating EPT1. *Biochem Biophys Res Commun.* (2019) 514:1251–6. doi: 10.1016/j.bbrc.2019.04.179

47. Wang HN, Chen YT, Mao XY, Du M. Maternal obesity impairs fetal mitochondriogenesis and brown adipose tissue development partially *via* upregulation of miR-204-5p. *Biochim Biophys Acta-Mol Basis Dis.* (2019) 1865:2706–15. doi: 10.1016/j.bbadis.2019.07.012

48. Park BS, Im HL, Yoon NA, Tu TH, Park JW, Kim JG, et al. Developmentally regulated GTP-binding protein-2 regulates adipocyte differentiation. *Biochem Biophys Res Commun.* (2021) 578:1–6. doi: 10.1016/j.bbrc.2021.08.081

49. Yang YY, Huang M, Wang YS. Targeted proteomic analysis of small GTPases in murine adipogenesis. *Anal Chem.* (2020) 92:6756–63. doi: 10.1021/acs.analchem.0c00974

50. Steensels S, Ersoy BA. Fatty acid activation in thermogenic adipose tissue. *Biochim Biophys Acta-Mol Cell Biol Lipids.* (2019) 1864:79–90. doi: 10.1016/j.bbalip.2018.05.008

51. Qi RL, Han X, Wang J, Qiu XY, Wang Q, Yang FY. MicroRNA-489-3p promotes adipogenesis by targeting the Postn gene in 3T3-L1 preadipocytes. *Life Sci.* (2021) 278:119620. doi: 10.1016/j.lfs.2021.119620

52. Tian N, Liu Q, Li Y, Tong L, Lu Y, Zhu Y, et al. Transketolase deficiency in adipose tissues protects mice from diet-induced obesity by promoting lipolysis. *Diabetes.* (2020) 69:1355–67. doi: 10.2337/db19-1087

53. Deng K, Ren C, Fan Y, Liu Z, Zhang G, Zhang Y, et al. miR-27a is an important adipogenesis regulator associated with differential lipid accumulation between intramuscular and subcutaneous adipose tissues of sheep. *Domest Anim Endocrinol.* (2020) 71:106393. doi: 10.1016/j.domaniend.2019.106393

54. Qiang J, Tao YF, He J, Sun YL, Xu P. miR-29a modulates SCD expression and is regulated in response to a saturated fatty acid diet in juvenile genetically improved farmed tilapia (*Oreochromis niloticus*). *J. Exp. Biol.* (2017) 220:1481–9. doi: 10.1242/jeb.151506

55. Yu HB, Zhao ZH, Yu XZ, Li JY, Lu CY, Yang RJ. Bovine lipid metabolism related gene GPAM: molecular characterization, function identification, and association analysis with fat deposition traits. *Gene.* (2017) 609:9–18. doi: 10.1016/j.gene.2017.01.031

frontiers | Frontiers in Genetics

TYPE Brief Research Report
PUBLISHED 06 January 2023
DOI 10.3389/fgene.2022.982433

OPEN ACCESS

EDITED BY
Lucas Lima Verardo,
Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Brazil

REVIEWED BY
Yahui Gao,
University of Maryland, United States
Xiangdong Ding,
China Agricultural University, China

*CORRESPONDENCE
Huijiang Gao,
gaohuijiang@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 30 June 2022
ACCEPTED 07 December 2022
PUBLISHED 06 January 2023

CITATION
Liang M, An B, Deng T, Du L, Li K, Cao S, Du Y, Xu L, Zhang L, Gao X, Cao Y, Zhao Y, Li J and Gao H (2023), Incorporating genome-wide and transcriptome-wide association studies to identify genetic elements of longissimus dorsi muscle in Huaxi cattle. *Front. Genet.* 13:982433. doi: 10.3389/fgene.2022.982433

Incorporating genome-wide and transcriptome-wide association studies to identify genetic elements of longissimus dorsi muscle in Huaxi cattle

Mang Liang[1†], Bingxing An[1†], Tianyu Deng[1], Lili Du[1], Keanning Li[1], Sheng Cao[1], Yueying Du[1], Lingyang Xu[1], Lupei Zhang[1], Xue Gao[1], Yang Cao[2], Yuming Zhao[2], Junya Li[1] and Huijiang Gao[1*]

[1]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Jilin Academy of Agricultural Sciences, Changchun, China

Locating the genetic variation of important livestock and poultry economic traits is essential for genetic improvement in breeding programs. Identifying the candidate genes for the productive ability of Huaxi cattle was one crucial element for practical breeding. Based on the genotype and phenotype data of 1,478 individuals and the RNA-seq data of 120 individuals contained in 1,478 individuals, we implemented genome-wide association studies (GWAS), transcriptome-wide association studies (TWAS), and Fisher's combined test (FCT) to identify the candidate genes for the carcass trait, the weight of longissimus dorsi muscle (LDM). The results indicated that GWAS, TWAS, and FCT identified seven candidate genes for LDM altogether: *PENK* was located by GWAS and FCT, *PPAT* was located by TWAS and FCT, and *XKR4*, *MTMR3*, *FGFRL1*, *DHRS4*, and *LAP3* were only located by one of the methods. After functional analysis of these candidate genes and referring to the reported studies, we found that they were mainly functional in the progress of the development of the body and the growth of muscle cells. Combining advanced breeding techniques such as gene editing with our study will significantly accelerate the genetic improvement for the future breeding of Huaxi cattle.

KEYWORDS

longissimus dorsi muscle, GWAS, TWAS, FCT, Huaxi cattle

Introduction

In ancient China, cattle, as the primary means of production, were mainly used as the draft ox and rarely considered the source of meat. With the rapid development of the economy, consumers' demand for beef, concerning quantity and quality, has increased in China. There is an urgent need to improve the productivity and quality of beef for the beef breed in China by directly changing the production capacity of beef cattle.

Locating the genetic variation of important livestock and poultry economic traits is still essential for genetic improvement. The genome-wide association study (GWAS) has successfully identified thousands of loci associated with complex features (Watanabe et al., 2019). However, 90% of the associated single nucleotide polymorphisms (SNPs) are located in the non-coding region of the gene, and their functions still are unknown, so the molecular mechanism of phenotypic variation cannot be explained clearly (Cannon and Mohlke, 2018). Previous studies have proved that gene expression is important in the phenotype of human diseases (He et al., 2013), and many genetic variations associated with phenotypes were likely to be expression quantitative trait loci (eQTL) (Nicolae et al., 2010). Furthermore, eQTL can be used to estimate the effects on gene expression and then be combined with physical phenotypes to conduct transcriptome-wide association studies (TWAS) to identify pivotal expression–trait associations (Gusev et al., 2016). The TWAS algorithm has been successfully implemented to identify the causal genes for the essential quantitative trait in cattle (Koupaie et al., 2019; Liu et al., 2021).

In this study, we utilized three strategies to identify the candidate genes that significantly affect the producibility of Huaxi cattle. First, we applied GWAS to identify the candidate gene by using 1,478 Huaxi cattle genotypes with the phenotypes of longissimus dorsi muscle (LDM) weight. Second, we implemented TWAS with genotypes (1,478 individuals), gene expression data of 120 individuals (contained in the 1,478 individuals), and phenotypes. Third, we utilized an ensemble approach, Fisher's test (Yu et al., 2008; Kremling et al., 2019), combining the results of GWAS and TWAS to identify the candidate gene. Finally, we analyzed the function and preliminarily explored the molecular mechanism of the candidate genes with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, which was helpful to the following breeding of Huaxi cattle.

Materials and methods

Animal resources and phenotype: The Huaxi cattle population, including 1,478 cattle born between 2008 and 2021, was established in Ulgai, Xilingol League, and Inner Mongolia of China. After weaning, all calves were moved to the Jinweifuren fattening farm in Beijing, where they shared uniform management and standardized feeding [they were fed with the total mixed ratio (TMR) according to the eighth revised edition of the Nutrition Requirements of Beef Cattle (NRC, 2006)]. Animals were slaughtered at 22–26 months of age with electrical stunning, followed by bloodletting. The weight of the longissimus dorsi muscle (LCM, kg) was weighed after being chilled at 4°C for 24 h.

Genotype and quality control: Genomic DNA was isolated from blood samples using the TIANamp Blood DNA Kit (Tiangen Biotech Co., Ltd., Beijing, China). DNA quality was acceptable when the A260/A280 ratio was in the range of 1.8–2.0. All individuals were genotyped using an Illumina BovineHD BeadChip that contained 770,000 SNPs. Quality control (QC) procedures were carried out using PLINK v1.9 (Purcell et al., 2007) to filter out SNPs with call rate <90%, minor allele frequency (MAF) < 0.05, and a significant deviation from the Hardy–Weinberg equilibrium ($p < 10^{-6}$), and >10% animals with missing genotype data were removed from the analysis. Finally, 1,478 cattle with 607,198 SNPs on 29 autosomal chromosomes with an average distance of 3 kb were included in subsequent analyses.

RNA extraction, library construction, sequencing, and quality control: Total RNA was extracted from SAT samples using TRIzol reagent (Invitrogen, Life Technologies) following the manufacturers' instructions. The RNA concentration, purity, and integrity were, respectively, analyzed on Qubit RNA Assay Kit (Life Technologies, CA, United States), NanoPhotometer Spectrophotometer (Thermo Fisher Scientific, MA, United States), and RNA Nano 6000 Assay Kit of the Bioanalyzer 2,100 system (Agilent Technologies, CA, United States). The high-quality samples with 28S/18S > 1.8 and OD 260/280 ratio >1.9 were applied for constructing cDNA libraries according to the protocol of IlluminaTruSeqTM RNA Kit (Illumina, United States). Samples that presented an RNA integrity number greater than 7.0 were then sent for paired-end RNA sequencing (read length 150 bp) on the Illumina NovaSeq 6,000 platform (Modi et al., 2021). The RNA sequencing was completed by Beijing Novogene Technology Co., Ltd. Trimmomatic (v0.39) was applied to remove the reads containing low-quality reads, poly-N, and adaptor sequences (Bolger et al., 2014). Sequentially, the clean reads were aligned to the *Bos taurus* reference genome ARS-UCD1.2 using HISAT2 (v2.2.1) (Lachmann et al., 2020), and then the generated SAM files were converted to BAM files through SAMtools (v1.11). featureCounts (v1.5.2) was used to estimate read counts (Liao et al., 2014).

GWAS: GWAS analysis of LDM traits based on the linear mixed model (LMM) was completed using GEMMA (Zhou and Stephens, 2012):

$$y = Xb + Sg + Z\alpha + e,$$

where y is the vector of phenotypes, b is the vector of fixed effect including age, sex, farm, and the days of fattening, S is the indicator variables of SNPs (0, 1, 2), g is the effect vector of SNPs, α is the polygenic effect vector, $\alpha \sim N(0, K\sigma_g^2)$, e is the random residual, and $e \sim N(0, I\sigma_e^2)$. In GWAS, the Wald test was used to test the SNP significance, and the threshold of the p-value was set at 1/m, where m is the number of SNPs (Wu et al., 2014).

TWAS: REML (restricted maximum likelihood) was utilized to evaluate the heritability of each gene base on the gene expression and cis-SNPs located within 1 Mb of the physical position of the gene. Then, the gene with significantly non-zero heritability will be incorporated in the subsequent analysis. For the preselected gene, Bayesian Sparse LMM (BSLMM) was used to estimate the effect values of the cis-SNPs for gene expression, and the prediction model that estimated gene expression with

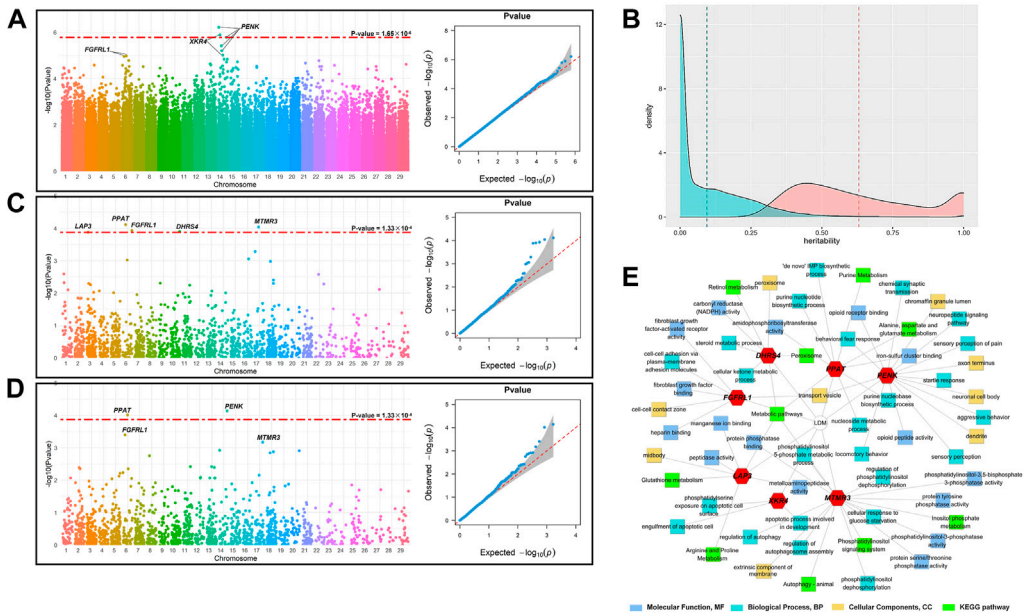


FIGURE 1
Identification of the candidate genes for LDM. **(A)** Manhattan plot and QQ plot of GWAS; the red dashed line indicates the threshold of Bonferroni's multiple test, $p = 1.65 \times 10^{-6}$. **(B)** Manhattan and QQ plots of TWAS; the red dashed line indicates the threshold of the corrected p-value with FDR = 0.01 ($p = 1.33 \times 10^{-4}$). **(C)** Manhattan and QQ plots of FCT; the red dashed line indicates the threshold of the corrected p-value with FDR = 0.01 ($p = 1.33 \times 10^{-4}$). **(D)** Distribution of the estimated heritability of the genes. The blue area represents the distribution of the heritability of all gene expression, and the blue dashed line represents the mean of the heritability estimates of converged gene expression (0.152 ± 0.263); the orange area represents the expression of 1,650 significant genes ($p < 0.05$), and the orange dashed line represents the mean of the heritability estimates of significant gene expression (0.631 ± 0.324). **(E)** Results of GO and KEGG analyses of the candidate genes.

cis-SNPs was constructed (Zhou et al., 2013). Afterward, the prediction model was utilized to estimate the gene expression values of the 1,358 individuals without transcriptome sequencing data but with genotypes (Dai et al., 2019; Zhou et al., 2020). Finally, all of the gene expression data were integrated with phenotypes to implement TWAS with LMM:

$$y = Xb + Wu + e,$$

where y and b are the same as in GWAS, W is the design matrix of the gene expression matrix, which is constructed with transcripts per kilobase million (TPM) (Luningham et al., 2020), u is the vector of gene effect, e is the random residual, and $e \sim N(0, I\sigma_e^2)$. In TWAS, the significant gene test was implemented with FDR, and the threshold of the p-value was set at $FDR \times n/m$, where FDR = 0.01, n is the number of genes with a p-value < 0.01, and m is the total number of genes in the LMM (Benjamini and Hochberg, 1995).

Fisher's combined test (FCT): The p-value in GWAS of each SNP in the top 10% of most associated SNPs was assigned to the nearest gene and then combined with the p-value in TWAS (linear model with multi-dimensional scaling (MDS) principal coordinates + 5 probabilistic estimation of expression residuals (PEERs)) for that same gene using Fisher's combined test as implemented in the sumlog method in the metap package (Dewey 2017) in R. TWAS p-values for genes which were not tested in TWAS was set to $p = 1$ prior to combining with GWAS p-values (Kremling et al., 2019). Similarly, the significant gene test was implemented in FCT using FDR with an identical threshold of the p-value.

Gene functional analysis: Gene Ontology (GO) is a database describing the function of genes and proteins. It annotated the genes into three types of terms: MF, BP, and CC (Ashburner et al., 2000). The KEGG database integrated the genome, regulatory network, and system function information (Kanehisa et al., 2016). To explore the function of candidate genes, we applied DAVID (https://david.ncifcrf.gov/) to implement GO and KEGG analyses of the genes and constructed the associated network of the gene-participated terms using ToppCluster (https://toppcluster.cchmc.org/).

Results

Genome-wide association studies

Figure 1A shows the Manhattan plot and QQ-plot of the GWAS analysis of LDM. The QQ-plot showed that there was no

TABLE 1 Details of the significantly associated SNPs identified by GWAS.

SNP	Chromosome	Location[a]	MAF[b]	Length[c]	Candidate gene[d]	*p*-value[e]
BovineHD1400006836	14	23,552,180	0.35	5,312	*PENK*	6.09E-07
BTB-00557532	14	24,643,266	0.38	32,311	*XKR4*	1.26E-06
BovineHD4100011289	14	23,553,712	0.22	6,844	*PENK*	1.63E-06

[a]The SNP position (bp) on ARS-UCD1.2.
[b]The minor allele frequency.
[c]The distance between SNP and the nearest gene.
[d]The nearest genes found on the Ensemble database (www.ensembl.org).
[e]*p*-values calculated by LMM.

TABLE 2 Details of the five candidate genes identified by TWAS.

Gene	Chr	Start[a]	End[a]	Effect ± SD[b]	*p*-value[c]	h^2 ± SD[d]
PPAT	6	71,782,614	71,821,764	-0.0031 ± 0.00089	7.68E-05	0.76 ± 0.32
MTMR3	17	68,971,211	69,102,722	-0.0060 ± 0.00020	9.11E-05	0.61 ± 0.31
FGFRL1	6	117,346,407	117,358,800	0.0020 ± 0.00068	1.17E-04	0.58 ± 0.30
DHRS4	10	21,088,232	21,100,627	-0.0069 ± 0.0022	1.26E-04	0.68 ± 0.32
LAP3	3	37,140,752	37,166,191	0.018 ± 0.0031	1.32E-04	0.70 ± 0.34

[a]The SNP position (bp) on ARS-UCD1.2.
[b]The effects of gene expression calculated by LMM in TWAS.
[c]*p*-values calculated by LMM.
[d]The heritability of gene expression calculated by REML.

apparent systematic deviation. Most of the points were distributed around the diagonal (the expansion coefficient is 1.05), which means that only a few SNPs were associated with the phenotype. The threshold of the *p*-value ($p = 1.65 \times 10^{-6}$) was set with Bonferroni's multiple test, and three SNPs in the 14th chromosome were significantly associated with the phenotypes, among which BovineHD1400006836 and BovineHD4100011289 were annotated to *PENK*, BTB-00557532 was annotated to *XKR4*, and the reference cattle genome was ARS-UCD1.2 more details are demonstrated in Table 1.

Transcriptome-wide association studies

After removing the genes with the average TPM (transcripts per kilobase million) less than 0.1, the expression levels of 15,325 genes of 120 individuals were assigned as phenotypes and 15,401 cis-SNPs, located within 1 Mb of the physical position of the gene, were assigned as the genotypes, and the heritability of the gene expression was estimated with REML. As shown in Figure 1B, the heritability of 15,324 genes converged in the progress of REML, and the average heritability was 0.152 ± 0.263. With $p < 0.05$ as the threshold, 1,650 genes were retained for the subsequent analysis, with an average heritability of 0.631 ± 0.324.

The Manhattan plot and QQ plot of TWAS for LDM were demonstrated in Figure 1C. The QQ-plot indicated that most points were distributed around the diagonal (expansion coefficient $\lambda = 1.03$), and several genes were significantly associated with LDM. After being corrected for the false discovery rate (FDR) of 0.01, the threshold of the *p*-value was set at 1.33×10^{-4}, and five genes were found to be significantly associated with LDM. The location of these genes is listed in Table 2. *PPAT* ($p = 7.68 \times 10^{-5}$), *MTMR3* ($p = 9.11 \times 10^{-5}$), *FGFRL1* ($p = 1.17 \times 10^{-4}$), *DHRS4* ($p = 1.26 \times 10^{-4}$), and *LAP3* ($p = 1.32 \times 10^{-4}$) were located in chromosomes 6, 17, 6, 10, and 3, respectively.

Fisher's combined test

The Manhattan plot and QQ plot of FCT analysis are shown in Figure 1D. The expansion coefficient λ of the QQ-plot was 1.02 with no systematic deviation, and most points were distributed on the diagonal, with only a minority of points floating above the diagonal. As with TWAS, the threshold of the *p*-value was set at 1.33×10^{-4}. The Manhattan plot indicated

that FCT identified two candidate genes significantly associated with LDM, namely, *PPAT* ($p = 9.69 \times 10^{-5}$) and *PENK* ($p = 7.26 \times 10^{-5}$), which were also identified by TWAS and GWAS, respectively.

Functional analysis of candidate genes

Combining the results of GWAS, TWAS, and FCT, *PENK*, *XKR4*, *PPAT*, *MTMR3*, *FGFRL1*, *DHRS4*, and *LAP3* were identified as the candidate genes of LDM. To further explore the function of these genes, we performed GO and KEGG analyses of these genes. The results are demonstrated in Figure 1E. These candidate genes participated in 48 GO terms, which contained 16 molecular function (MF) terms, 23 biological progress (BP) terms, and nine cellular component (CC) terms. For MF, the candidate genes mainly function in the progress of fibroblast growth factor activity regulation (GO:0005007 and GO:0017134), NADPH activity (GO:0004090), and serine, threonine, and tyrosine metabolism (GO:0004722 and GO:0004725). KEGG pathway analysis found that candidate genes were involved in 10 pathways, mainly including amino acid and peptide metabolism, signal transduction pathway, purine metabolism, and other biological processes.

Discussion

Abundant studies have proven that GWAS could precisely locate the candidate loci for the quantitative traits in livestock breeding, especially for the traits with high heritability. It was one of the most widespread methods used in plant and animal improvement programs. However, the regulatory mechanism from SNP to phenotypic variation was still unknown in most cases, and it was impossible to determine the genuine pathogenic gene of the trait associated with the candidate SNPs due to the linkage disequilibrium (LD) in the SNPs. In recent years, the innovation of sequencing technology provided more other omics biological information, transcriptome, metabolome, *etc*., and assisted in locating candidate genes more accurately. TWAS implement the association analysis based on the gene expression data with the phenotype to locate the candidate genes directly. The results of previous studies indicated that TWAS performed well in practice (Dai et al., 2019; Luningham et al., 2020; Li et al., 2021). In this study, we not only performed GWAS and TWAS individually but also utilized an ensemble approach, FCT, combining the results of GWAS and TWAS to locate the candidate genes for LDM.

For LDM in this study, we indented seven candidate genes by GWAS, TWAS, and FCT in total: *PENK* was located by GWAS and FCT, *PPAT* was located by TWAS and FCT, and the remaining five genes were only located by one of the methods. An et al. (2019) also located *PENK*, which was associated with the height of Brahman cattle and Nerol cattle populations. The studies on humans also found that *PENK* regulated cell development by encoding the opioid peptide growth factor (ORF) to affect height (Pryce et al., 2011). Zhan et al. (2014) found that a variation site (8p12.1) in *XKR4* was associated with human thyroid-stimulating hormone (TSH) secretion, and it was the candidate gene for the development traits in Brahman cattle, Korean yellow cattle, Chinese Holstein cattle, and Chinese Sujiang pig populations (Lindholm-Perry et al., 2012; Edea et al., 2020; Naserkheil et al., 2020; Xu et al., 2020). The protein encoded by *PPAT* was a member of the purine/pyrimidine phosphoribosyltransferase family, which was essential in regulating the proliferation, migration, and invasion of thyroid cancer. Gene function analysis found *PPAT* functionals in inosinic acid biosynthesis (GO:0006189), and GART was the functional partner of *PPAT*, which had a fundamental impact on nucleotide metabolism and internal environment balance (Welin et al., 2010). *MTMR3* is a member of the MTM family associated with muscular dysplasia, which participates in the cell progress of proliferation, differentiation, autophagy, and division by regulating the synthesis of myotube (Hnia et al., 2012). The reported studies have confirmed that *MTMR3* was the virtually candidate gene in the Holstein population for the quantitative traits, such as milk fat rate, milk yield, and milk protein content (Pimentel et al., 2011). *FGFRL1* encoded fibroblast growth factor receptor one, which plays a crucial role in the progress of cell adhesion, embryonic slow muscle fiber development, and bone tissue formation (Amann et al., 2014; Niu et al., 2015; Yang et al., 2016). Bluteau et al. (2014) indicated a slight reduction in the whole bone of the *FGFRL1* gene knockout mice. The study on Holstein also identified *FGFRL1* as a candidate gene for development traits in the Holstein population (Zhang et al., 2017). *DHRS4* encodes NADP(H)-dependent retinol dehydrogenase/reductase. The study on pigs found that rs196958886, one of the SNPs of this gene, may induce the peroxisome proliferator-activated receptor alpha (PPARα) gene, affect the interaction between fatty acids and glucose metabolism, and ultimately affect the quality of pork (Hwang et al., 2017). *LAP3* encodes leucine aminopeptidase, which is functional in protein metabolism and growth (Yao et al., 2020). Substantial studies on cattle found that *LAP3* was a candidate gene that affects important production traits such as visceral organ weight, body size, and carcass traits (Setoguchi et al., 2009; Bongiorni et al., 2012; Xia et al., 2017; An et al., 2018; An et al., 2020). Zheng et al. (2011) implemented association analyses between *LAP3* and milking traits in the Holstein population and concluded that *LAP3* was the vital candidate gene for milking traits.

Conclusion

In conclusion, we identified seven candidate genes of LDM by GWAS, TWAS, and FCT based on genome and transcriptome information. According to the previous relevant studies and the results of gene function analysis, the candidate genes were mainly functional in the progress of the development of the body and the growth of muscle cells. Combining advanced breeding techniques such as gene editing with our study will significantly accelerate the genetic improvement of Huaxi cattle.

Data availability statement

The data presented in the study are deposited in the NCBI repository (https://www.ncbi.nlm.nih.gov/), accession number PRJNA721166, and DRYAD repository (https://datadryad.org/stash), accession number 10.5061/dryad.4qc06.

Author contributions

ML wrote the paper. JL and HG revised the paper. BA and LD performed experiments. TD, KL, SC, and YD collected the data. YC and YZ provided fund support. LX, XG, and LZ participated in the design of the study and contributed to the acquisition of data. All authors reviewed and approved the final manuscript.

Funding

This work was supported by funds from the National Natural Science Foundation of China (32172693) and the Program of National Beef Cattle and Yak Industrial Technology System (CARS-37). The China Agriculture Research System of MOF and MARA supported the statistical analysis and writing of the paper. The Technology Project of Inner Mongolia Autonomous Region (2020GG0210) and the Open project of key laboratory of Beef Cattle Genetics and Breeding, Ministry of Agriculture and Rural Affairs P.R.China (BC2022kfkt01) also supported this work.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amann, R., Wyder, S., Slavotinek, A. M., and Trueb, B. (2014). The FgfrL1 receptor is required for development of slow muscle fibers. *Dev. Biol.* 394 (2), 228–241. doi:10.1016/j.ydbio.2014.08.016

An, B., Xia, J., Chang, T., Wang, X., Miao, J., Xu, L., et al. (2018). Genome-wide association study identifies loci and candidate genes for internal organ weights in Simmental beef cattle. *Physiol. Genomics* 50 (7), 523–531. doi:10.1152/physiolgenomics.00022.2018

An, B., Xia, J., Chang, T., Wang, X., Xu, L., Zhang, L., et al. (2019). Genome-wide association study reveals candidate genes associated with body measurement traits in Chinese Wagyu beef cattle. *Anim. Genet.* 50 (4), 386–390. doi:10.1111/age.12805

An, B., Xu, L., Xia, J., Wang, X., Miao, J., Chang, T., et al. (2020). Multiple association analysis of loci and candidate genes that regulate body size at three growth stages in Simmental beef cattle. *BMC Genet.* 21 (1), 32–11. doi:10.1186/s12863-020-0837-6

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

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 Methodol.* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bluteau, G., Zhuang, L., Amann, R., and Trueb, B. (2014). Targeted disruption of the intracellular domain of receptor FgfrL1 in mice. *PloS one* 9 (8), e105210. doi:10.1371/journal.pone.0105210

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170

Bongiorni, S., Mancini, G., Chillemi, G., Pariset, L., and Valentini, A. (2012). Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. *PLoS One* 7 (12), e50137. doi:10.1371/journal.pone.0050137

Cannon, M. E., and Mohlke, K. L. (2018). Deciphering the emerging complexities of molecular mechanisms at GWAS loci. *Am. J. Hum. Genet.* 103 (5), 637–653. doi:10.1016/j.ajhg.2018.10.001

Dai, Y., Pei, G., Zhao, Z., and Jia, P. (2019). A convergent study of genetic variants associated with crohn's disease: Evidence from GWAS, gene expression, methylation, eQTL and TWAS. *Front. Genet.* 10, 318. doi:10.3389/fgene.2019.00318

Dewey, M. (2017). Metap: Meta-analysis of significance values r-package.

Edea, Z., Jung, K. S., Shin, S.-S., Yoo, S.-W., Choi, J. W., and Kim, K.-S. (2020). Signatures of positive selection underlying beef production traits in Korean cattle breeds. *J. Anim. Sci. Technol.* 62 (3), 293–305. doi:10.5187/jast.2020.62.3.293

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi:10.1038/ng.3506

He, X., Fuller, C. K., Song, Y., Meng, Q., Zhang, B., Yang, X., et al. (2013). Sherlock: Detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.* 92 (5), 667–680. doi:10.1016/j.ajhg.2013.03.022

Hnia, K., Vaccari, I., Bolino, A., and Laporte, J. (2012). Myotubularin phosphoinositide phosphatases: Cellular functions and disease pathophysiology. *Trends Mol. Med.* 18 (6), 317–327. doi:10.1016/j.molmed.2012.04.004

Hwang, J. H., An, S. M., Kwon, S. G., Park, D. H., Kim, T. W., Kang, D. G., et al. (2017). Associations of the polymorphisms in DHRS4, SERPING1, and APOR genes with postmortem pH in berkshire pigs. *Anim. Biotechnol.* 28 (4), 288–293. doi:10.1080/10495398.2017.1279171

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 (1), D457–D462. doi:10.1093/nar/gkv1070

Koupaie, E. H., Azizi, A., Lakeh, A. B., Hafez, H., and Elbeshbishy, E. (2019). A comprehensive dataset on anaerobic digestion of cattle manure, source separated organics, and municipal sludge using different inoculum sources. *Data Brief.* 24, 103913. doi:10.1016/j.dib.2019.103913

Kremling, K. A., Diepenbrock, C. H., Gore, M. A., Buckler, E. S., and Bandillo, N. B. (2019). Transcriptome-wide association supplements genome-wide association in Zea mays. *G3* 9 (9), 3023–3033. doi:10.1534/g3.119.400549

Lachmann, A., Clarke, D. J., Torre, D., Xie, Z., and Ma'ayan, A. (2020). Interoperable RNA-Seq analysis in the cloud. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1863 (6), 194521. doi:10.1016/j.bbagrm.2020.194521

Li, D., Liu, Q., and Schnable, P. S. (2021). TWAS results are complementary to and less affected by linkage disequilibrium than GWAS. *Plant Physiol.* 186 (4), 1800–1811. doi:10.1093/plphys/kiab161

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

Lindholm-Perry, A., Kuehn, L., Smith, T., Ferrell, C., Jenkins, T., Freetly, H., et al. (2012). A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle 1. *Anim. Genet.* 43 (2), 216–219. doi:10.1111/j.1365-2052.2011.02232.x

Liu, D., Chen, Z., Zhao, W., Guo, L., Sun, H., Zhu, K., et al. (2021). Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. *BMC genomics* 22 (1), 1–19. doi:10.1186/s12864-021-08042-x

Luningham, J. M., Chen, J., Tang, S., De Jager, P. L., Bennett, D. A., Buchman, A. S., et al. (2020). Bayesian genome-wide TWAS method to leverage both cis-and trans-eQTL information through summary statistics. *Am. J. Hum. Genet.* 107 (4), 714–726. doi:10.1016/j.ajhg.2020.08.022

Modi, A., Vai, S., Caramelli, D., and Lari, M. (2021). "The Illumina sequencing protocol and the NovaSeq 6000 system," in *Bacterial pangenomics* (New York, United States: Springer), 15–42.

Naserkheil, M., Bahrami, A., Lee, D., and Mehrban, H. (2020). Integrating single-step GWAS and bipartite networks reconstruction provides novel insights into yearling weight and carcass traits in hanwoo beef cattle. *Animals.* 10 (10), 1836. doi:10.3390/ani10101836

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet.* 6 (4), e1000888. doi:10.1371/journal.pgen.1000888

Niu, T., Liu, N., Zhao, M., Xie, G., Zhang, L., Li, J., et al. (2015). Identification of a novel FGFRL1 MicroRNA target site polymorphism for bone mineral density in meta-analyses of genome-wide association studies. *Hum. Mol. Genet.* 24 (16), 4710–4727. doi:10.1093/hmg/ddv144

NRC (2006). "NY/T 1180-2006. Determination of meat tenderness and shear force method ," in *Beijing: Ministry of Agricultre of the PRC.*

Pimentel, E., Bauersachs, S., Tietze, M., Simianer, H., Tetens, J., Thaller, G., et al. (2011). Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Anim. Genet.* 42 (3), 251–262. doi:10.1111/j.1365-2052.2010.02148.x

Pryce, J. E., Hayes, B. J., Bolormaa, S., and Goddard, M. E. (2011). Polymorphic regions affecting human height also control stature in cattle. *Genetics* 187 (3), 981–984. doi:10.1534/genetics.110.123943

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

Setoguchi, K., Furuta, M., Hirano, T., Nagao, T., Watanabe, T., Sugimoto, Y., et al. (2009). Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC Genet.* 10 (1), 43–12. doi:10.1186/1471-2156-10-43

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J., et al. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* 51 (9), 1339–1348. doi:10.1038/s41588-019-0481-0

Welin, M., Grossmann, J. G., Flodin, S., Nyman, T., Stenmark, P., Tresaugues, L., et al. (2010). Structural studies of tri-functional human GART. *Nucleic Acids Res.* 38 (20), 7308–7319. doi:10.1093/nar/gkq595

Wu, Y., Fan, H., Wang, Y., Zhang, L., Gao, X., Chen, Y., et al. (2014). Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. *PloS one* 9 (10), e109330. doi:10.1371/journal.pone.0109330

Xia, J., Fan, H., Chang, T., Xu, L., Zhang, W., Song, Y., et al. (2017). Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. *Sci. Rep.* 7 (1), 42048–42049. doi:10.1038/srep42048

Xu, P., Ni, L., Tao, Y., Ma, Z., Hu, T., Zhao, X., et al. (2020). Genome-wide association study for growth and fatness traits in Chinese Sujiang pigs. *Anim. Genet.* 51 (2), 314–318. doi:10.1111/age.12899

Yang, X., Steinberg, F., Zhuang, L., Bessey, R., and Trueb, B. (2016). Receptor FGFRL1 does not promote cell proliferation but induces cell adhesion. *Int. J. Mol. Med.* 38 (1), 30–38. doi:10.3892/ijmm.2016.2601

Yao, H., Liu, C., Lin, D., Liu, S., Lin, Z., and Dong, Y. (2020). Polymorphisms of LAP3 gene and their association with the growth traits in the razor clam Sinonovacula constricta. *Mol. Biol. Rep.* 47 (2), 1257–1264. doi:10.1007/s11033-019-05231-6

Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* 178 (1), 539–551. doi:10.1534/genetics.107.074245

Zhan, M., Chen, G., Pan, C.-M., Gu, Z.-H., Zhao, S.-X., Liu, W., et al. (2014). Genome-wide association study identifies a novel susceptibility gene for serum TSH levels in Chinese populations. *Hum. Mol. Genet.* 23 (20), 5505–5517. doi:10.1093/hmg/ddu250

Zhang, X., Chu, Q., Guo, G., Dong, G., Li, X., Zhang, Q., et al. (2017). Genome-wide association studies identified multiple genetic loci for body size at four growth stages in Chinese Holstein cattle. *PloS one* 12 (4), e0175971. doi:10.1371/journal.pone.0175971

Zheng, X., Ju, Z., Wang, J., Li, Q., Huang, J., Zhang, A., et al. (2011). Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. *Mol. Biol. Rep.* 38 (6), 4053–4061. doi:10.1007/s11033-010-0524-1

Zhou, D., Jiang, Y., Zhong, X., Cox, N. J., Liu, C., and Gamazon, E. R. (2020). A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat. Genet.* 52 (11), 1239–1246. doi:10.1038/s41588-020-0706-2

Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* 9 (2), e1003264. doi:10.1371/journal.pgen.1003264

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44 (7), 821–824. doi:10.1038/ng.2310

frontiers | Frontiers in Genetics

TYPE Data Report
PUBLISHED 24 February 2023
DOI 10.3389/fgene.2023.1038606

OPEN ACCESS

EDITED BY
Nuno Carolino,
Instituto Nacional Investigaciao Agraria e Veterinaria (INIAV), Portugal

REVIEWED BY
Juan Vicente Delgado Bermejo,
Juan Vicente Delgado bermejo, Spain
Elena Ciani,
University of Bari Aldo Moro, Italy

*CORRESPONDENCE
Hao Sun,
sunhao92@jlu.edu.cn
Xueqi Sun,
bc_lky@163.com

†These authors have contributed equally to this work and share first authorship

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

RECEIVED 07 September 2022
ACCEPTED 20 January 2023
PUBLISHED 24 February 2023

CITATION
Ni H, Zhang Y, Yang Y, Yin Y, Xie H, Zheng J, Dong L, Diao J, Wei M, Lv Z, Yan S, Li Y, Sun H and Sun X (2023), Whole genome sequencing of a wild swan goose population.
Front. Genet. 14:1038606.
doi: 10.3389/fgene.2023.1038606

Whole genome sequencing of a wild swan goose population

Hongyu Ni[1†], Yonghong Zhang[1†], Yuwei Yang[1], Yijing Yin[1], Hengli Xie[1], Jinlei Zheng[1], Liping Dong[1], Jizhe Diao[1], Meng Wei[1], Zhichao Lv[2], Shouqing Yan[1], Yumei Li[1], Hao Sun[1*] and Xueqi Sun[1,3*]

[1]College of Animal Science, Jilin University, Changchun, China, [2]Jilin Province Economic Management Cadre College, Changchun, China, [3]Jilin Academy of Agricultural Sciences, Changchun, China

KEYWORDS

swan goose, whole genome sequencing (WGS), SNP, linkage disequilibrium, genetic resource

Introduction

It was reported that Chinese goose breeds are thought to have originated from wild swan geese (*Anser cygnoides*) (Shi et al., 2006). The wild swan goose, as a migratory waterfowl, has many different characteristics compared to the domestic goose breeds. Most notable is that the wild swan goose has long-distance flight ability. At present, the swan goose has two biological migration routes, namely inland migration routes and coastal migration routes. The majority of swan goose individuals in northeastern Mongolia (Inner Mongolia, Heilongjiang, and Jilin province) relocated to the Chinese Yangtze River, and the Russian Far East relocated to the coast of southeast China for winter (Zhu et al., 2020). Due to the lack of analysis of the genomic characteristics and selection signals of wild swan geese, the genetic basis underlying these characteristics is still not well investigated.

The swan goose has been listed as vulnerable in the Red List of Threatened Species of the International Union for Conservation of Nature (IUCN), and the swan goose's global population has been estimated at c.60,000–90,000 (http://www.iucnredlist.org/). In China, the wild swan goose has been designated as a Class II national protected species. Therefore, it's difficult to obtain wild swan goose samples to carry out genomic studies. Xianghai wetland national nature reserve (44°55′–45°09′N, 122°05′–122°31′E), located in the Jilin province (the northeast region of China), was one of the wild swan goose's habitats. In 1981, the Xianghai wetland national nature reserve was established, and it is the habitat of many protected migratory birds e.g. *Grus japonensis, Otis tarda*.

We are working at the front line of geese breeding in the Xianghai wetland national nature reserve. Our team consists of researchers from Jilin University and Jiuzhou Flying Goose Husbandry&Technology Co., Ltd., which holds a license to domesticate swan geese in the Xianghai Wetlands. In 1999, their farm was established for domesticating and breeding wild swan geese in Yanya lake in the Xianghai wetland national nature reserve. In recent years, we are working on hybrid goose breeding (swan goose × local domestic goose). We are interested in investigating the genetic difference between the wild swan goose and the domestic goose. Considering that sequencing the wild swan goose provides a valuable resource not only for researchers working on animal conservation but also for those focusing on goose genomic breeding and further genomic investigation. With this in mind, we report and make publicly available whole genome sequencing data for 10 wild swan geese.

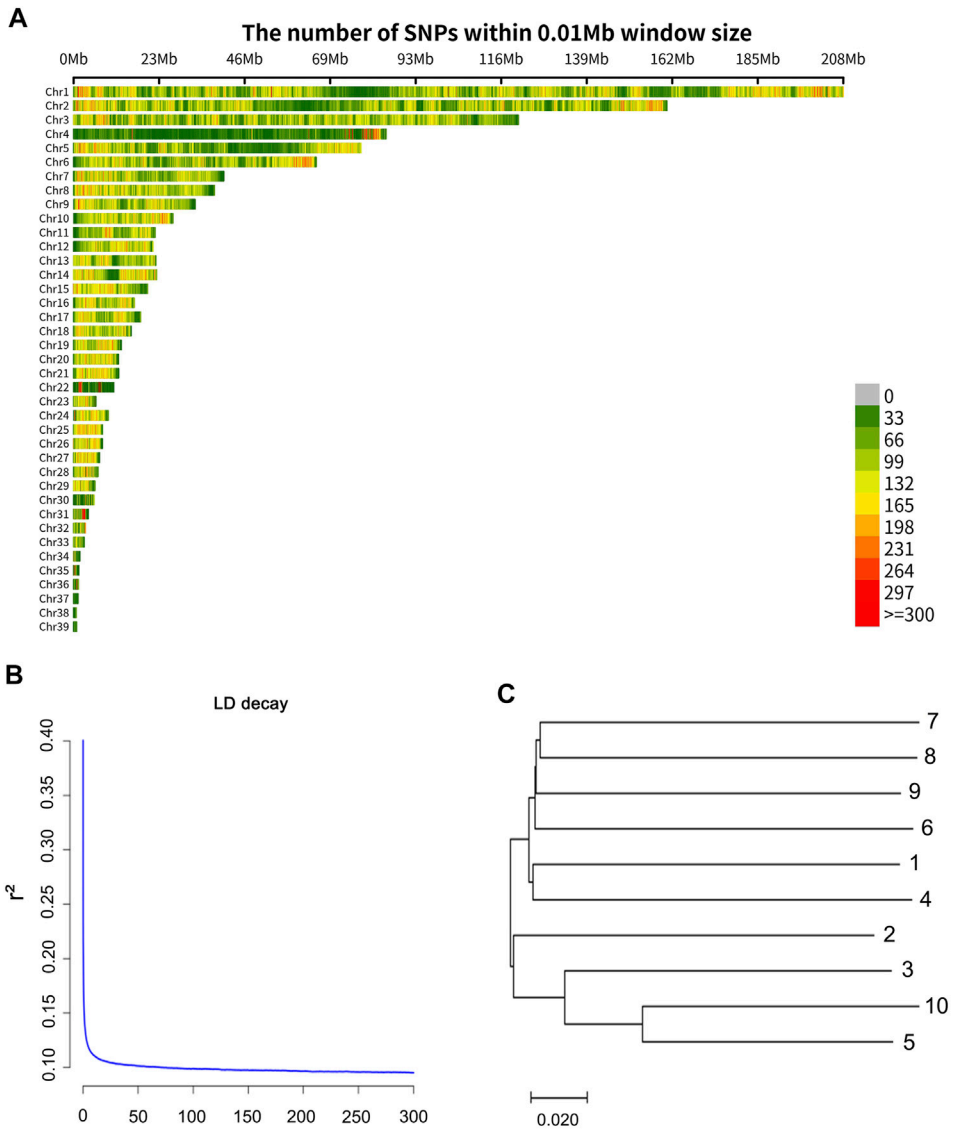


FIGURE 1
(A) Distribution of the SNPs on the chromosomes. The *x*-axis represents the chromosome position (Mb), and the *y*-axis represents the chromosomes. The number of SNPs presented in each 10 kb genome block is displayed by the different colors. **(B)** Extent of LD average r^2 values at distances up to 300 kb. **(C)** The neighbor-joining tree based on 1-IBS distance.

Samples collection and sequencing

The Jiuzhou Flying Goose Husbandry&Technology Co., Ltd. holds a license to domesticate swan geese in the Xianghai Wetlands. The experienced staff randomly picked up wild swan goose eggs from the Xianghai Wetland National Nature Reserve for incubation and rearing. About 2 mL of blood samples were collected from the veins under the wings of the adult swan geese by the experienced staff, and all the swan geese remained healthy after blood collection. Genomic DNA was extracted from the blood following the standard phenol-chloroform extraction procedure. For genome sequencing, at least 0.5 μg of genomic DNA from each sample was used to construct a library with an insert size of 350 bp. Paired-end sequencing libraries were constructed according to the manufacturer's instructions (Illumina Inc., San Diego, CA, United States) and sequenced on the Illumina HiSeq platform.

Data quality control and variant calling

The FASTP (Chen et al., 2018) software was used to perform quality control on the raw data. The clean reads were aligned to the swan goose genome (PRJNA826973) using Burrows-Wheeler Alignment Maximal Extract Matches algorithm (Li and Durbin,

2009) with default parameters. The SNPs (Single Nucleotide Polymorphisms) were called using the GATK software (McKenna et al., 2010). For these to be called, the calling quality had to be greater than 20 (base recognition accuracy >99%). The SNPs were filtered with minor allele frequency <0.05 and missing rate >0.10 using the VCFtools (Danecek et al., 2011). The distribution of the SNPs on the chromosomes was plotted using the R "CMplot" package.

Data description

A total of 162.6 Gbp clean data was obtained. The sequence data were deposited in the NCBI Sequence Read Archive (SRA) and the accession number of the sequencing data is PRJNA814334. One individual was deeply sequenced with 39.6 Gbp data obtained, and the other samples were sequenced with about 13.7 Gbp data obtained. The sequencing information of each sample is shown in Supplementary Table S1. The average mapping rate was 0.90 (mapping quality value ≥20). Finally, a total of 10,727,005 SNPs (minor allele frequency ≥0.05) were obtained, and a total of 7,646,295 (71.28%) transitions (Ts) and 3,080,710 (28.72%) transversions (Tv) were observed. The distribution of the SNPs in 10 kb non-overlapping windows on the chromosomes is provided in Figure 1A. From the information of the SNPs, Linkage disequilibrium (r2) measures were calculated based on the PopLDdecay (Zhang et al., 2019). The decay of LD (linkage disequilibrium) according to distance, for pair-wise SNPs up to 300 kb is shown in Figure 1B. The LD value (r^2 = 0.2) was about 200 bp in the wild swan goose, and the levels of LD at different distances were presented in Supplementary Table S2.

To provide future researchers with an understanding of the genetic difference between the wild swan goose and the domestic goose, here we compared the swan goose populations with a domestic goose breed, the Zi goose, a local domesticated breed in northeast China. The PCA and LD analysis results are provided in Supplementary Figure S1. The plot of the first two PCA components showed a clear genetic differentiation, along the first component, between the swan geese and Zi geese. In addition, the PCA patterning highlighted a wider genetic differentiation among the swan geese samples than among the Zi geese (Supplementary Figure S1A). Historically, the wild swan geese have not undergone intensive selection as experienced in Zi geese. Hence, it is reasonable to assume that the Swan geese mantain higher levels of genetic variability than Zi geese. This assumption is also supported by the LD analysis. In our study, we find that the LD extend of the Zi geese is higher than the Swan geese (Supplementary Figure S1B). It is known that the LD extent could reflect the history of the population. The strong artificial selection could contribute to increasing the LD extent, although we can't exclude the possibility of other factors such as gene flow.

To provide future researchers with an understanding of the genetic relationship among samples, we constructed an inter-individual neighbor-joining (NJ) tree of the ten swan geese samples (Figure 1C) using the (1-IBS) genetic distances, with the identity-by-state (IBS) values being estimated using PLINK v1.9 (Purcell et al., 2007). Genetic distances between individuals ranged from 0.184 to 0.291, and the average value was 0.272. Moreover, we calculated the inbreeding coefficient of each individual using PLINK v1.9 (Purcell et al., 2007) with the command "het." The inbreeding coefficient (Supplementary Table S3) ranged from −0.01 to 0.12, and the average value was 0.07. This result indicated that the extent of inbreeding of the wild swan goose population is low.

In conclusion, this paper provides the whole genome resequencing data of ten swan geese, which can serve as a theoretical reference for future exploration of genetic variation or characteristics between swan geese and domestic goose populations, and is of great value in revealing the genetic mechanism of phenotypic differences. For example, by comparing the genome of the wild swan goose to the local goose, researchers may find the clue to the genetic background of the special characteristics of the swan goose, e.g., long-term flying ability.

Data availability statement

The sequence data presented in the study are deposited in the NCBI Sequence Read Achieve (SRA) repository, accession number PRJNA814334. The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of Jilin University.

Author contributions

YZ and HS designed the research. YZ, XS, YL, SY, ZL, JD, and LD collected the samples. HN, YuY, YiY, HX, MW, and JZ performed the experiment. HS and HN analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

Funding

This study was supported by the Key Projects of Jilin Province Science and Technology Development Plan (Grant No. 20220203084SF).

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1038606/full#supplementary-material

References

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

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

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Shi, X. W., Wang, J. W., Zeng, F. T., and Qiu, X. P. (2006). Mitochondrial DNA cleavage patterns distinguish independent origin of Chinese domestic geese and Western domestic geese. *Biochem. Genet.* 44, 237–245. doi:10.1007/s10528-006-9028-z

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi:10.1093/bioinformatics/bty875

Zhu, Q., Hobson, K. A., Zhao, Q. S., Zhou, Y. Q., Damba, I., Batbayar, N., et al. (2020). Migratory connectivity of Swan Geese based on species' distribution models, feather stable isotope assignment and satellite tracking. *Divers. Distributions* 26 (8), 944–957. doi:10.1111/ddi.13077

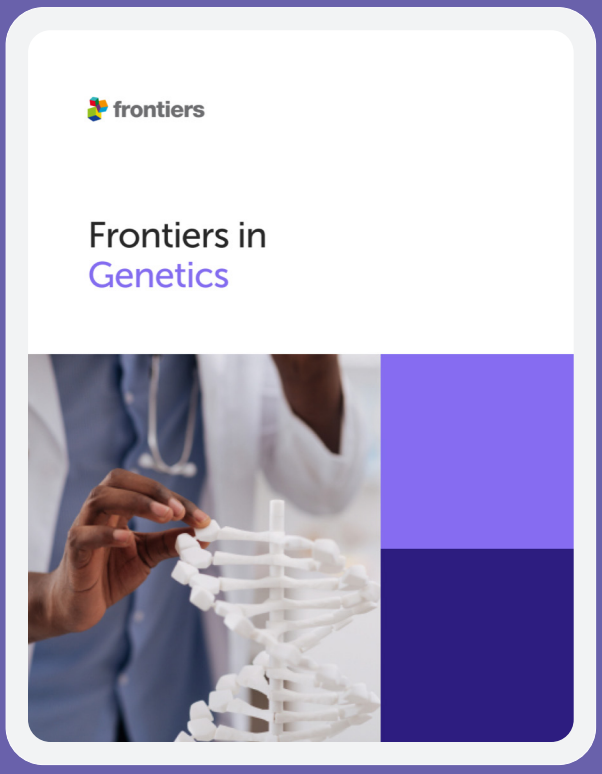
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact