frontiers | Research Topics

Genetic markers identification for animal production and disease resistance

Edited by
Hongyu Liu, Ibrar Muhammad Khan, Muhammad Zahoor Khan and Adnan Khan

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-3096-2
DOI 10.3389/978-2-8325-3096-2

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

Genetic markers identification for animal production and disease resistance

Topic editors

Hongyu Liu — Anhui Agricultural University, China

Ibrar Muhammad Khan — Anhui Agricultural University, China

Muhammad Zahoor Khan — University of Agriculture, Dera Ismail Khan, Pakistan

Adnan Khan — Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China

Citation

Liu, H., Khan, I. M., Khan, M. Z., Khan, A., eds. (2023). *Genetic markers identification for animal production and disease resistance*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3096-2

Table of contents

05 Editorial: Genetic markers identification for animal production and disease resistance
Ibrar Muhammad Khan, Adnan Khan, Hongyu Liu and Muhammad Zahoor Khan

08 Genetic polymorphisms of *PKLR* gene and their associations with milk production traits in Chinese Holstein cows
Aixia Du, Fengru Zhao, Yanan Liu, Lingna Xu, Kewei Chen, Dongxiao Sun and Bo Han

24 Transcriptional analysis of microRNAs related to unsaturated fatty acid synthesis by interfering bovine adipocyte ACSL1 gene
Xupeng Li, Yanbin Bai, Jingsheng Li, Zongchang Chen, Yong Ma, Bingang Shi, Xiangmin Han, Yuzhu Luo, Jiang Hu, Jiqing Wang, Xiu Liu, Shaobin Li and Zhidong Zhao

35 Genetic architecture and selection of Anhui autochthonous pig population revealed by whole genome resequencing
Wei Zhang, Xiaojin Li, Yao Jiang, Mei Zhou, Linqing Liu, Shiguang Su, Chengliang Xu, Xueting Li and Chonglong Wang

48 A genome-wide integrated analysis of lncRNA-mRNA in melanocytes from white and brown skin hair boer goats (*Capra aegagrus hircus*)
Ji Kai-yuan, Zhao Yi-Wei, Wen Ru-jun, Ibrar Muhammad Khan and Zhang Yun-hai

59 miRNA profiling in intrauterine exosomes of pregnant cattle on day 7
Yaying Zhai, Qiaoting Shi, Qiuxia Chu, Fuying Chen, Yajie Feng, Zijing Zhang, Xinglei Qi, Danny Arends, Gudrun A. Brockmann, Eryao Wang and Shijie Lyu

68 Genetic marker identification of *SEC13* gene for milk production traits in Chinese holstein
Ruike Jia, Lingna Xu, Dongxiao Sun and Bo Han

78 Unraveling signatures of chicken genetic diversity and divergent selection in breed-specific patterns of early myogenesis, nitric oxide metabolism and post-hatch growth
Ivan I. Kochish, Vladimir Yu. Titov, Ilya N. Nikonov, Evgeni A. Brazhnik, Nikolai I. Vorobyov, Maxim V. Korenyuga, Olga V. Myasnikova, Anna M. Dolgorukova, Darren K. Griffin and Michael N. Romanov

95 Genetic association of wool quality characteristics in United States Rambouillet sheep
Gabrielle M. Becker, Julia L. Woods, Christopher S. Schauer, Whit C. Stewart and Brenda M. Murdoch

108 Population structure, genetic diversity and prolificacy in pishan red sheep under an extreme desert environment
Cheng-long Zhang, Jihu Zhang, Mirenisa Tuersuntuoheti, Qianqian Chang and Shudong Liu

119 The ovine *HIAT1* gene: mRNA expression, InDel mutations, and growth trait associations
Yunyun Luo, Zhanerke Akhatayeva, Cui Mao, Fugui Jiang, Zhengang Guo, Hongwei Xu and Xianyong Lan

128 Exploring the association between fat-related traits in chickens and the RGS16 gene: insights from polymorphism and functional validation analysis
Mao Ye, Zhexia Fan, Yuhang Xu, Kang Luan, Lijin Guo, Siyu Zhang and Qingbin Luo

139 Transcriptome analysis to identify candidate genes related to mammary gland development of Bactrian camel (*Camelus bactrianus*)
Huaibing Yao, Xiaorui Liang, Zhihua Dou, Zhongkai Zhao, Wanpeng Ma, Zelin Hao, Hui Yan, Yuzhuo Wang, Zhuangyuan Wu, Gangliang Chen and Jie Yang

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 12 July 2023
DOI 10.3389/fgene.2023.1243793

Check for updates

OPEN ACCESS

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

*CORRESPONDENCE
Hongyu Liu,
liuhongyu@ahau.edu.cn
Muhammad Zahoor Khan,
zahoorkhattak91@yahoo.com

†These authors have contributed equally to this work

RECEIVED 21 June 2023
ACCEPTED 03 July 2023
PUBLISHED 12 July 2023

CITATION
Khan IM, Khan A, Liu H and Khan MZ (2023), Editorial: Genetic markers identification for animal production and disease resistance.
Front. Genet. 14:1243793.
doi: 10.3389/fgene.2023.1243793

COPYRIGHT
© 2023 Khan, Khan, Liu and Khan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Genetic markers identification for animal production and disease resistance

Ibrar Muhammad Khan[1†], Adnan Khan[2†], Hongyu Liu[3*] and Muhammad Zahoor Khan[4*]

[1]Anhui Province Key Laboratory of Embryo Development and Reproduction Regulation, Anhui Province Key Laboratory of Environmental Hormone and Reproduction, School of Biological and Food Engineering, Fuyang Normal University, Fuyang, China, [2]Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [3]Anhui Provincial Laboratory of Local Livestock and Poultry Genetical Resource Conservation and Breeding, College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [4]Department of Animal Breeding and Genetics, Faculty of Veterinary and Animal Sciences, The University of Agriculture, Dera Ismail Khan, Pakistan

KEYWORDS

production traits, genomics, genetic markers, animal health, markers assisted selection

Editorial on the Research Topic
Genetic markers identification for animal production and disease resistance

It is well established that several production traits and diseases are polygenetic in nature (van Rheenen et al., 2019). Traditional methods have proven inadequate in effectively enhancing production performance and disease control in animals. Thus, identification of markers and the underlying mechanism for controlling these phenotypes is the main focus of research today in animal science (Goddard and Hayes, 2009). Various genetic approaches including RNA-sequencing (Augustino et al., 2020; Khan et al., 2020; Liu et al., 2022), whole-genome sequencing, mapping of quantitative trait loci (QTL) (Uemoto et al., 2021), candidate gene analysis (Yang et al., 2016) and genome-wide association study (GWAS) (Kai-Yuan et al.) (Wang et al., 2022) have been utilized to identify fundamental genes or their polymorphisms correlated with animal production and disease resistance phenotypic traits in animals (Ma et al., 2021; Khan et al., 2023). The utilization of markers for milk quality, production traits, disease resistance (Khan et al., 2022; Yang et al., 2022), thermo-tolerance, fertility, and carcass quality in cattle plays a crucial role in enhancing their health and productivity (Khan et al., 2021). These markers enable targeted selection and breeding programs, optimizing production and ensuring reduce infection risks in animals.

Li et al. investigated the role of ACSL1 in fatty acid synthesis and metabolism, specifically focusing on its impact on highly unsaturated fatty acids in beef. They transfected bovine preadipocytes with si-ACSL1 and NC-ACSL1, and used RNA-Seq to identify miRNAs associated with unsaturated fatty acid synthesis. By analyzing the miRNA-mRNA interaction network, several key miRNA-mRNA targeting relationships were identified, including novel-m0035-5p—ACSL1, novel-m0035-5p—ELOVL4, miR-9-x—ACSL1, bta-miR-677—ACSL1, miR-129-x—ELOVL4, and bta-miR-485—FADS2. These specific miRNAs have the potential to regulate unsaturated fatty acid synthesis in bovine adipocytes by targeting these specific genes. A study conducted by Du et al. utilized resequencing techniques to identify 21 single nucleotide polymorphisms (SNPs)

in the PKLR gene of Chinese Holstein cows. Their analysis demonstrated significant associations between these SNPs and milk yield, fat and protein yields, as well as protein percentage. Additionally, in a separate study by Jia et al., two specific SNPs (SNP g.54362761A>G and g.54326411T>C) in the SEC13 gene were documented. Notably, Jia et al. observed a significant association between these SNPs and milk production phenotypic traits in dairy cattle. Another study by Zhai et al. revealed that intrauterine exosomes in bovine pregnancy play a role in embryo development and implantation. Analysis of miRNA expression in these exosomes revealed significant differences between donor and recipient cows, with 22 miRNAs upregulated and 38 downregulated in the donor group. The identified miRNAs were associated with biological processes and pathways related to embryo implantation and endometrial development, suggesting their potential influence on the uterine microenvironment for successful implantation of embryos.

Luo et al. examined HIAT1 expression in Lanzhou fat-tailed (LFT) sheep and found it widely expressed, with high levels in the testis. They also discovered a 9-bp insertion mutation (rs1089950828) in Luxi black-headed (LXBH) and Guiqian semi-fine wool (GSFW) sheep breeds. The mutation was associated with morphometric traits in LXBH and GSFW sheep. Yearling ewes with a heterozygous genotype (ID) had smaller body sizes, while yearling rams and adult ewes showed better growth performance. Furthermore, the authors suggested the potential use of this mutation for marker-assisted selection in Chinese sheep populations.

Fine wool production significantly contributes to the revenue of extensive wool sheep production in the United States. In their study, Becker et al. focused on Rambouillet sheep, renowned for their high-quality wool, aiming to unravel the genetic basis of wool characteristics. Over a 3-year period, wool samples and DNA were collected from rams participating in performance tests. Genome-wide association studies revealed significant genetic markers linked to wool quality traits, specifically on chromosomes 1, 2, 4, 15, and 19. The study also identified strong correlations between various wool characteristics, such as clean fleece weight and staple length. Moreover, certain wool traits exhibited associations with body weight and scrotal circumference. These findings provide valuable insights for Rambouillet sheep breeders, aiding them in making informed decisions regarding targeted selection and breeding to enhance wool quality.

Climate changes, particularly extreme weather events, can have a negative impact on livestock production. To understand the genetic mechanisms behind sheep prolificacy traits in the Taklimakan Desert, Zhang et al. conducted a study using Pishan Red Sheep (PRS) and Qira Black Sheep (QR). Blood samples were collected from these sheep, and DNA was extracted and analyzed using the Illumina Ovine SNP50 chip. The study revealed the genetic characteristics of PRS, including linkage disequilibrium, effective population size, and genetic markers identified through iHS and FST analyses. A total of 29 genes were found to be common in both analyses. These findings provide valuable insights for the conservation of sheep genetic resources and molecular breeding in desert environments.

Kai-Yuan et al. explored the roles of long noncoding RNAs (lncRNAs) in melanocytes, which are responsible for determining skin and hair color through melanin production. Melanocytes from Boer goat skins were isolated and characterized using various staining and immunohistochemical techniques. A phenotypic analysis revealed significant differences in the behavior and melanin production of melanocytes from goats with white and brown hair. RNA sequencing was performed, leading to the identification of candidate lncRNAs and mRNAs involved in stage-specific melanogenesis. Functional enrichment analysis highlighted the involvement of miRNA precursors and cisregulatory effects of lncRNAs. The study also proposed multiple lncRNA-mRNA networks related to melanocyte migration, proliferation, and melanogenesis, offering new insights into mammalian pigmentation.

A recent experimental trial conducted by Zhang et al. aimed to investigate the genetic diversity and population structure of pigs in Anhui Province, China. The study involved the sequencing of 150 pig genomes from six representative populations, which were then compared to data from Asian wild boars and commercial pigs. The analysis uncovered two distinct ancestral origins of Anhui pigs, namely, the Wannan Spotted pig (WSP) and Wannan Black pig (WBP) from a common ancestor, while the remaining four populations originated from a separate ancestor. The researchers also identified specific genomic regions associated with domestication traits in Anhui pigs, such as reproduction, lipid and meat characteristics, and ear size genes (CABS1, INSL6, MAP3K12, IGF1R, INSR, LIMK2, PATZ1, MAPK1, SNX19, MSTN, MC5R, PRKG1, CREBBP, and ADCY9). These findings not only enhance our understanding of genetic variations among pig populations but also hold promise for future genetic research and the advancement of genome-assisted breeding in pigs and other domesticated animals.

The study conducted by Yao et al. aimed to identify candidate genes associated with mammary gland development in Bactrian camels (*Camelus bactrianus*) through transcriptome analysis. The researchers analyzed the gene expression profiles of the mammary gland tissues and identified differentially expressed genes (DEGs) involved in the development of the mammary gland. Functional enrichment analysis provided insights into the biological processes, molecular functions, and cellular components associated with mammary gland development. Several candidate genes, including those involved in epithelial development, mammary gland morphogenesis, and milk synthesis, were identified. This research contributes to a better understanding of the genetic mechanisms underlying mammary gland development in Bactrian camels, which may have implications for improving camel milk production and reproductive performance.

Kochish et al. investigated the genetic diversity and divergent selection in chickens, focusing on breed-specific patterns related to early myogenesis, nitric oxide metabolism, and post-hatch growth. The researchers analyzed the genetic signatures of various chicken breeds and identified differentially expressed genes (DEGs) associated with these traits. They observed breed-specific patterns in early myogenesis, suggesting unique genetic pathways involved in muscle development across different chicken breeds. Additionally, variations in nitric oxide metabolism and post-hatch growth were found among the breeds. This study sheds light on the genetic basis of these traits in chickens, providing valuable insights for future

breeding strategies aimed at enhancing muscle development and growth in poultry. In addition, Ye et al. investigated the association between fat-related traits in chickens and the RGS16 gene. They examined polymorphisms within the RGS16 gene and performed functional validation analysis to gain insights into its role in fat-related traits. The study aimed to understand the potential impact of genetic variations in the RGS16 gene on fat deposition and metabolism in chickens. The findings provide valuable information on the genetic factors influencing fat-related traits in poultry and contribute to a better understanding of the molecular mechanisms underlying fat deposition and metabolism. This research may have implications for improving breeding strategies and the production of leaner and healthier chicken meat.

Author contributions

IK and MK initiated the Research Topic and drafted the manuscript. AK, IK, MK, and HL co-edited the Research Topic and participated in the editorial process. All authors contributed to the article and approved the submitted version.

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

Augustino, S. M., Xu, Q., Liu, X., Liu, L., Zhang, Q., and Yu, Y. (2020). Transcriptomic study of porcine small intestine epithelial cells reveals important genes and pathways associated with susceptibility to *Escherichia coli* F4ac diarrhea. *Front. Genet.* 11, 68. doi:10.3389/fgene.2020.00068

Goddard, M., and Hayes, B. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi:10.1038/nrg2575

Khan, A., Dou, J., Wang, Y., Jiang, X., Khan, M. Z., Luo, H., et al. (2020). Evaluation of heat stress effects on cellular and transcriptional adaptation of bovine granulosa cells. *J. animal Sci. Biotechnol.* 11 (1), 25–20. doi:10.1186/s40104-019-0408-8

Khan, M. Z., Dari, G., Khan, A., and Yu, Y. (2022). Genetic polymorphisms of TRAPPC9 and CD4 genes and their association with milk production and mastitis resistance phenotypic traits in Chinese Holstein. *Front. Veterinary Sci.* 9, 1008497. doi:10.3389/fvets.2022.1008497

Khan, M. Z., Ma, Y., Ma, J., Xiao, J., Liu, Y., Liu, S., et al. (2021). Association of DGAT1 with cattle, buffalo, goat, and sheep milk and meat production traits. *Front. Veterinary Sci.* 8, 712470. doi:10.3389/fvets.2021.712470

Khan, M. Z., Wang, J., Ma, Y., Chen, T., Ma, M., Ullah, Q., et al. (2023). Genetic polymorphisms in immune-and inflammation-associated genes and their association with bovine mastitis resistance/susceptibility. *Front. Immunol.* 14, 1082144. doi:10.3389/fimmu.2023.1082144

Liu, H., Khan, I. M., Liu, Y., Khan, N. M., Ji, K., Yin, H., et al. (2022). A comprehensive sequencing analysis of testis-born miRNAs in immature and mature indigenous wandong cattle (*Bos taurus*). *Genes* 13 (12), 2185. doi:10.3390/genes13122185

Ma, Y., Khan, M. Z., Xiao, J., Alugongo, G. M., Chen, X., Chen, T., et al. (2021). Genetic markers associated with milk production traits in dairy cattle. *Agriculture* 11 (10), 1018. doi:10.3390/agriculture11101018

Uemoto, Y., Ichinoseki, K., Matsumoto, T., Oka, N., Takamori, H., Kadowaki, H., et al. (2021). Genome-wide association studies for production, respiratory disease, and immune-related traits in Landrace pigs. *Sci. Rep.* 11 (1), 15823. doi:10.1038/s41598-021-95339-2

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., and Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nat. Rev. Genet.* 20, 567–581. doi:10.1038/s41576-019-0137-z

Wang, P., Li, X., Zhu, Y., Wei, J., Zhang, C., Kong, Q., et al. (2022). Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. *Front. Veterinary Sci.* 9, 932034. doi:10.3389/fvets.2022.932034

Yang, J., Tang, Y., Liu, X., Zhang, J., Zahoor Khan, M., Mi, S., et al. (2022). Characterization of peripheral white blood cells transcriptome to unravel the regulatory signatures of bovine subclinical mastitis resistance. *Front. Genet.* 13, 949850. doi:10.3389/fgene.2022.949850

Yang, S., Gao, Y., Zhang, S., Zhang, Q., and Sun, D. (2016). Identification of genetic associations and functional polymorphisms of SAA1 gene affecting milk production traits in dairy cattle. *Plos one* 11 (9), 2195. doi:10.1371/journal.pone.0162195

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fgene.2022.1002706

Check for updates

OPEN ACCESS

EDITED BY
Ibrar Muhammad Khan,
Fuyang Normal University, China

REVIEWED BY
Muhammad Zahoor Khan,
University of Agriculture, Dera Ismail Khan, Pakistan
Kaiyuan Ji,
Anhui Agricultural University, China

*CORRESPONDENCE
Bo Han,
bohan@cau.edu.cn/hanbo_98@126.com

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

RECEIVED 25 July 2022
ACCEPTED 12 August 2022
PUBLISHED 02 September 2022

CITATION
Du A, Zhao F, Liu Y, Xu L, Chen K, Sun D and Han B (2022), Genetic polymorphisms of PKLR gene and their associations with milk production traits in Chinese Holstein cows. *Front. Genet.* 13:1002706. doi: 10.3389/fgene.2022.1002706

COPYRIGHT
© 2022 Du, Zhao, Liu, Xu, Chen, Sun and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic polymorphisms of *PKLR* gene and their associations with milk production traits in Chinese Holstein cows

Aixia Du[1], Fengru Zhao[2], Yanan Liu[1], Lingna Xu[1], Kewei Chen[3], Dongxiao Sun[1] and Bo Han[1]*

[1]National Engineering Laboratory of Animal Breeding, Key Laboratory of Animal Genetics, Department of Animal Genetics and Breeding, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, China Agricultural University, Beijing, China, [2]Beijing Dairy Cattle Center, Beijing, China, [3]Yantai Institute, China Agricultural University, Yantai, China

Our previous work had confirmed that pyruvate kinase L/R (*PKLR*) gene was expressed differently in different lactation periods of dairy cattle, and participated in lipid metabolism through insulin, PI3K-Akt, MAPK, AMPK, mTOR, and PPAR signaling pathways, suggesting that *PKLR* is a candidate gene to affect milk production traits in dairy cattle. Here, we verified whether this gene has significant genetic association with milk yield and composition traits in a Chinese Holstein cow population. In total, we identified 21 single nucleotide polymorphisms (SNPs) by resequencing the entire coding region and partial flanking region of *PKLR* gene, in which, two SNPs were located in 5′ promoter region, two in 5′ untranslated region (UTR), three in introns, five in exons, six in 3′ UTR and three in 3′ flanking region. The single marker association analysis displayed that all SNPs were significantly associated with milk yield, fat and protein yields or protein percentage ($p \leq 0.0497$). The haplotype block containing all the SNPs, predicted by Haploview, had a significant association with fat yield and protein percentage ($p \leq 0.0145$). Further, four SNPs in 5′ regulatory region and eight SNPs in UTR and exon regions were predicted to change the transcription factor binding sites (TFBSs) and mRNA secondary structure, respectively, thus affecting the expression of *PKLR*, leading to changes in milk production phenotypes, suggesting that these SNPs might be the potential functional mutations for milk production traits in dairy cattle. In conclusion, we demonstrated that *PKLR* had significant genetic effects on milk production traits, and the SNPs with significant genetic effects could be used as candidate genetic markers for genomic selection (GS) in dairy cattle.

KEYWORDS

PKLR, milk production traits, association analysis, GS, SNP chips

Introduction

Milk is rich in nutrition and is an important food for the human body to obtain many essential nutrients. Fat and protein in milk have the characteristics of easy digestion and absorption, especially for children and the elderly, so the content and proportion of fat and protein in milk is of great significance. Studies have shown that drinking milk can reduce the incidence of dental caries (Rumbold et al., 2021), cardiovascular disease (Soedamah-Muthu and de Goede 2018), metabolic syndrome (Crichton et al., 2011) and obesity (Abargouei et al., 2012). Dairy cattle breeding is essential for the development of the dairy industry and human health. In dairy cattle breeding, one of the most important thing is to study the milk production traits, milk yield, fat yield, and percentage, and protein yield and percentage, which are quantitative traits and controlled by multiple minor polygenes, a few main efficient genes and greatly affected by the environment (Schrooten et al., 2000). However, the process of traditional breeding is very slow and unable to meet the growing consumer demand.

Meuwissen et al. (2001) first proposed genomic selection (GS) in 2001, which can better reflect the problem of minorgenes for quantitative traits (Wiggans et al., 2011). Especially for animals such as dairy cattle with long generation interval, GS can effectively shorten their generation interval and accelerate genetic progress (Stock and Reents 2013). Since 2009, GS has been formally applied to dairy cattle breeding, which has brought revolutionary changes to dairy cattle breeding (Wiggans et al., 2017). SNP (single nucleotide polymorphism) chips designed with SNP probes based on large-scale SNP genotype data to detect genomic polymorphism (Heffner et al., 2009) were used in GS to select target traits. In recent years, with the development of SNP chip technology, GS has been widely used in dairy cattle breeding (Jiang et al., 2013; Jiang et al., 2016). Through GS, a single marker whose effect is small can be captured (Goddard and Hayes 2007). Additionally, studies have shown that adding functional site information with large genetic effects on target traits can improve the accuracy of GS (Zhang et al., 2014; Brondum et al., 2015; Zhang et al., 2015; de Las Heras-Saldana et al., 2020). Therefore, in recent years, researchers have been using various methods such as quantitative trait locus (QTL) mapping, candidate gene analysis, genome-wide association study (GWAS) and high throughput omics strategy to explore functional genes and mutations related to milk production traits, so as to improve the accuracy of GS and accelerate the process of molecular breeding of dairy cattle (Gebreyesus et al., 2019; Lopdell et al., 2019; Liu et al., 2020; Korkuc et al., 2021). At present, in terms of milk producing traits of dairy cattle, many genes such as *CDKN1A*, *FADS2*, *PRLR*, *SLC2A12,* and *SLC5A1* had been verified to be associated with milk yield and composition traits of Holstein cows (Maryam et al., 2015; Han et al., 2017; Yan et al., 2018; Shi et al., 2019; Valsalan et al., 2021; Zwierzchowski et al., 2021; Fu et al., 2022).

Previously, we obtained liver transcriptome data of Chinese Holstein cows at different lactations, and found that pyruvate kinase L/R (*PKLR*) gene was differentially expressed during periods and participated in lipid metabolism through insulin, PI3K-Akt, MAPK, AMPK, mTOR, and PPAR signaling pathways, suggesting that *PKLR* gene may play an important role for milk fat trait of dairy cattle (Liang et al., 2017). *PKLR* is involved in glycogen and lipid metabolisms in liver tissues (Wang et al., 2000; Ahrens et al., 2013), and has a wide association with a spectrum of liver damage from steatosis and inflammation to fibrosis *via* its regulation on mitochondrial dysfunction and subsequent hepatic triglyceride accumulation (Chella Krishnan et al., 2021). In addition, *PKLR* (chr.3: 15344765-15354042) is located 0.02 Mb to the peak of QTL regions for milk fat percentage (QTL_ID: 104486) and protein percentage (QTL_ID:104816, 104938) (Nayeri et al., 2016). Therefore, we considered this gene to be a potential candidate gene for milk producing traits in dairy cows.

Herein, we identified SNPs of the *PKLR* gene in a Chinese Holstein population and analyzed their genetic associations with milk yield, fat yield, fat percentage, protein yield and protein percentage. Further, we predicted the potential biological effects of identified SNPs on transcription factor binding site (TFBS) and mRNA secondary structure. The purpose of this study is to provide valuable SNP loci information for dairy GS, and also to provide some reference information for the in-depth study of the mechanism of candidate genes related to milk production traits in dairy cattle.

Materials and methods

Animals and phenotypic data

In this study, we used a total of 925 Chinese Holstein cows from 44 sire families for association analyses, and these cows were distributed in 21 dairy farms belonging to the Beijing Shounong Animal Husbandry Development Co., Ltd. (Beijing, China), where the cows were healthy with the same feeding conditions and had accurate pedigree information and standard dairy herd improvement (DHI) records. We used the phenotypic data of 925 cows in the first lactation and 633 in the second lactation (292 cows merely completed the milking of first lactation) for the association analyses and mainly analyzed five milk production traits, including 305-days milk yield, fat yield, fat percentage, protein yield and protein percentage. The descriptive statistics of phenotypic values for dairy production traits of the first and second lactations were presented in Supplementary Table S1.

DNA extraction

The Beijing Dairy Cattle Center (Beijing, China) provides frozen semen of the 44 bulls and blood samples of 925 cows that were stored at −20°C for genomic DNA extraction. We extracted frozen semen DNAs by salt-out procedure, and extracted DNAs of blood samples by a TIANamp Blood DNA Kit (Tiangen, Beijing, China). Then, we used NanoDrop 2000 Spectrophotometer (Thermo Scientific, Hudson, NH, United States) and the gel electrophoresis to determine the quantity and quality of the extracted DNAs, respectively.

SNP identification and genotyping

According to the sequences of bovine *PKLR* gene (NC_037330) from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), we used Primers3 (https://primer3.ut.ee/) to design the primers (Supplementary Table S2) in this gene's coding region, parts of intron region and 2,000 bp of upstream and downstream regions. The primers were synthesized by Beijing Genomics Institute (BGI, Beijing, China). We mixed the semen DNAs equally, amplified them by PCR (Supplementary Table S3), and detected the PCR amplification products using 2% gel electrophoresis before Sanger sequencing by BGI. After sequencing, we identified the potential SNPs according to the reference sequences (ARS-UCD1.2) on NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Subsequently, we genotyped the identified SNPs in 925 cows using Genotyping by Target Sequencing (GBTS) technology by Boruidi Biotechnology Co., Ltd. (Hebei, China).

Linkage disequilibrium estimation and association analyses

We used Haploview4.2 (Broad Institute of MIT and Harvard, Cambridge, MA, United States) to estimate the extent of linkage disequilibrium (LD) between the identified SNPs.

The MIXED process in SAS 9.4 (SAS Institute Inc., Cary, NC, United States) software was used to carry out association analyses between the genotypes/haplotype blocks and the five milk production traits, milk yield, fat yield, fat percentage, protein yield, and protein percentage, on the first and second lactations. The following animal model was used for the association analysis: $y = \mu + HYS + b \times M + G + a + e$; where y is the phenotypic value of each trait for each cow; μ is the overall mean; HYS is the fixed effect of farm (1–21 for 21 farms, respectively), year (1–4 for the year 2012–2015, respectively), and season (1 for April–May; 2 for June–August; 3 for September–November; and 4 for December– March); M is the age of calving as a covariant, b is the regression coefficient of covariant M; G is the genotype or haplotype combination effect; a is the individual random additive genetic effect, distributed as $N(0, A\delta_a^2)$, with the additive genetic variance δ_a^2; and e is the random residual, distributed as $N(0, I\delta_e^2)$, with identity matrix I and residual error variance δ_e^2.

Additionally, we calculated the additive effect (a), dominant effect (d), and substitution effect (α) by the following formulas: $a = \frac{AA-BB}{2}$, $d = AB - \frac{AA+BB}{2}$, $\alpha = a + d(q - p)$, where AA, BB and AB are the least square means of the milk production traits in the corresponding genotypes, p is the frequency of allele A, and q is the frequency of allele B.

Functional prediction of mutation sites

We predicted changes of TFBSs for the SNPs located in the 5′ region of *PKLR* gene by the MEME Suite (http://meme-suite.org/). We used RNAfold Web Server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) to predict changes in secondary structures of mRNA for SNPs in UTR and exon regions. The minimum free energy (MFE) of the optimal secondary structure reflects the stability of mRNA structure. The lower the MFE value, the more stable the mRNA structure is.

Results

SNPs identification

In this study, we totally found 21 SNPs in *PKLR* gene, all of which had been reported previously. Two SNPs, 3:g.15342877C>T and 3:g.15344349A>C, were located in 5′ promoter region, two (3:g.15345216C>T and 3:g.15345227T>C) in 5′ untranslated region (UTR), three (3:g.15349740A>G, 3:g.15350548C>T and 3:g.15350805T>C) in introns, five (3:g.15349768A>G, 3:g.15349978A>G, 3:g.15350655A>G, 3:g.15350898T>C and 3:g.15352855T>C) in exons, six (3:g.15353088A>C, 3:g.15353235T>C, 3:g.15353254T>C, 3:g.15353292C>G, 3:g.15353330A>G and 3:g.15353342C>T) in 3′ UTR, and three (3:g.15355389T>C, 3:g.15355514T>C and 3:g.15355833A>G) in 3′ flanking region. All the five SNPs in the exons were synonymous mutations (Table 1). The genotypic and allelic frequencies of all the identified SNPs were summarized in Table 1.

Associations between SNPs and five milk productions traits

We analyzed the associations between the 21 SNPs in *PKLR* and five milk production traits in dairy cattle. In the first lactation, there were four, nineteen, four and seventeen SNPs significantly associated with milk yield, fat yield, protein yield and protein percentage, respectively ($p \leq 0.0497$; Table 2). Four SNPs, 3:g.15350898T>C, 3:g.15355389T>C, 3:g.15355514T>C

TABLE 1 Details of SNPs identified in *PKLR* gene.

SNP name	GenBank no.	Location	Genotype	Genotypic frequency	Allele	Allelic frequency
3:g.15342877C>T	rs134381383	5′ promoter region	CC	0.0724	C	0.2876
			CT	0.4303	T	0.7124
			TT	0.4973		
3:g.15344349A>C	rs135669860	5′ promoter region	AA	0.0724	A	0.287
			AC	0.4292	C	0.713
			CC	0.4984		
3:g.15345216C>T	rs134794841	5′ UTR	CC	0.0724	C	0.287
			CT	0.4292	T	0.713
			TT	0.4984		
3:g.15345227T>C	rs110280638	5′ UTR	CC	0.4995	C	0.7135
			CT	0.4281	T	0.2865
			TT	0.0724		
3:g.15349740A>G	rs109049992	intron	AA	0.0714	A	0.2865
			AG	0.4303	G	0.7135
			GG	0.4984		
3:g.15349768A>G	rs110522117	exon 7	AA	0.0714	A	0.2865
			AG	0.4303	G	0.7135
			GG	0.4984		
3:g.15349978A>G	rs109620290	exon 7	AA	0.0714	A	0.2859
			AG	0.4292	G	0.7141
			GG	0.4994		
3:g.15350548C>T	rs109009333	intron	CC	0.0714	C	0.2865
			CT	0.4303	T	0.7135
			TT	0.4984		
3:g.15350655A>G	rs135555311	exon 9	AA	0.0714	A	0.2865
			AG	0.4303	G	0.7135
			GG	0.4984		
3:g.15350805T>C	rs109578013	intron	CC	0.4984	C	0.7135
			CT	0.4303	T	0.2865
			TT	0.0714		
3:g.15350898T>C	rs208110429	exon 10	CC	0.0281	C	0.1827
			CT	0.3092	T	0.8173
			TT	0.6627		
3:g.15352855T>C	rs109938041	exon 12	CC	0.4984	C	0.7135
			CT	0.4303	T	0.2865
			TT	0.0714		
3:g.15353088A>C	rs135526735	3′ UTR	AA	0.0714	A	0.287
			AC	0.4313	C	0.713
			CC	0.4973		
3:g.15353235T>C	rs109536098	3′ UTR	CC	0.4951	C	0.7114
			CT	0.4324	T	0.2886
			TT	0.0724		
3:g.15353254T>C	rs110474872	3′ UTR	CC	0.4951	C	0.7114
			CT	0.4324	T	0.2886
			TT	0.0724		
3:g.15353292C>G	rs136694042	3′ UTR	CC	0.0757	C	0.2908
			CG	0.4303	G	0.7092
			GG	0.494		

(Continued on following page)

TABLE 1 (*Continued*) Details of SNPs identified in *PKLR* gene.

SNP name	GenBank no.	Location	Genotype	Genotypic frequency	Allele	Allelic frequency
3:g.15353330A>G	rs135489031	3′ UTR	AA	0.0854	A	0.2957
			AG	0.4205	G	0.7043
			GG	0.4941		
3:g.15353342C>T	rs133320650	3′ UTR	CC	0.0886	C	0.2978
			CT	0.4184	T	0.7022
			TT	0.493		
3:g.15355389T>C	rs133757664	3′ flanking region	CC	0.4935	C	0.6876
			CT	0.3883	T	0.3124
			TT	0.1182		
3:g.15355514T>C	rs132659643	3′ flanking region	CC	0.4951	C	0.6891
			CT	0.3879	T	0.3109
			TT	0.117		
3:g.15355833A>G	rs108993332	3′ flanking region	AA	0.1159	A	0.3099
			AG	0.3879	G	0.6901
			GG	0.4962		

Note: UTR: untranslated region.

and 3:g.15355833A>G, had extremely significant genetic effects on milk, fat and protein yields ($p \leq 0.0044$), and 3:g.15355389T>C and 3:g.15355514T>C were also significantly associated with protein percentage ($p \leq 0.0374$). As for the second lactation, there were sixteen, twenty and eighteen SNPs were significantly associated with milk yield, fat yield and protein percentage ($p \leq 0.0436$), respectively. Additionally, thirteen SNPs were significantly associated with milk yield, fat yield and protein percentage ($p \leq 0.0063$). During two lactation periods, six SNPs, 3:g.15353292C>G, 3:g.15353330A>G, 3:g.15353342C>T, 3:g.15355389T>C, 3:g.15355514T>C and 3:g.15355833A>G, had significant genetic effects on fat yield ($p \leq 0.0097$). In addition, the results of allelic additive, dominant and substitution effects of the SNPs in *PKLR* gene were displayed in Supplementary Table S4.

Associations between haplotype block and five milk productions traits

We estimated the degree of linkage disequilibrium (LD) among the 21 identified SNPs in *PKLR* gene using Haploview4.2, and inferred one haplotype block including all the SNPs (Figure 1). The block consisted of four haplotypes, H1 (TCTCGGGTGCTCCCCGGTCCG), H2 (CACTAAACATTTATTCACTTA), H3 (TCTCGGGTGCCCCCCGGTCCG), and H4 (TCTCGGGTGCTCCCCGGTTTA) with the frequencies of 0.499, 0.287, 0.181, and 0.021, respectively. The haplotype combinations demonstrated significant associations with fat yield and protein percentage in the first and second lactations ($p \leq 0.0145$), and milk yield ($p = 0.0003$) and protein yield ($p = 0.0183$) in the second lactation (Supplementary Table S5).

Regulation of the 5′ region SNPs on transcriptional activity

We used the MEME Suite software to predict the changes of TFBSs caused by the four SNPs on the 5′ regulatory region of *PKLR* gene. The detailed results were shown in Table 3. The allele C of 3:g.15342877C>T created binding sites (BSs) for transcription factors (TFs) SP100 and ESRRA. In 3:g.15344349A>C, allele A created BSs for three TFs, MLX, ZBTB33 and IRF5, and the allele C created the BSs for ZNF524, YY2, and SREBF2. As for 3:g.15345216C>T, the allele C invented BS for RREB1, the allele T invented BSs for TWIST2, ZEB1, NAC007, BHLHE22, ZFP42, TCF3, NAC031, ZSCAN31, and TCF12. The allele C of 3:g.15345227T>C created BSs for TFs MYC, TFAP2A and TCF4.

Prediction of changes in secondary structures of mRNA

We used the RNAfold Web Server to predict the changes of secondary structures of mRNA for thirteen SNPs in UTR and exon regions of *PKLR* gene. All the thirteen SNP mutation sites were predicted to change the MFE of mRNA secondary structures compared to the MFE of reference sequence (XM_024989616.1; ARS-UCD1.2; Table 4). Among them, six sites, 3:g.15345216T, 3:g.15345227C, 3:g.15349768G, 3:g.15350898C, 3:g.15353235C and 3:g.15353254C, could

TABLE 2 Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
3: g.15342877C>T	1	CC (67)	9,994.06 ± 192.05	325.52 ± 8.0346^{AB}	3.2752 ± 0.07772	299.96 ± 5.857	3.0133 ± 0.02668^{a}
		CT (398)	10014 ± 179.5	327.21 ± 7.597^{A}	3.2885 ± 0.07293	297.22 ± 5.5367	2.9789 ± 0.02461^{b}
		TT (460)	9,970.53 ± 177.07	322.44 ± 7.5119^{B}	3.2569 ± 0.072	295.75 ± 5.4744	2.9785 ± 0.02422^{b}
		p	0.6498	0.0308	0.2369	0.1995	0.0313
	2	CC (43)	11507 ± 239.38^{A}	420.86 ± 10.0507^{A}	3.6252 ± 0.09705	332.26 ± 7.3264^{a}	2.8773 ± 0.03295^{A}
		CT (270)	11115 ± 221.72^{B}	413.29 ± 9.437^{A}	3.6831 ± 0.09033	325.59 ± 6.8773^{b}	2.93 ± 0.02998^{B}
		TT (320)	11064 ± 218.52^{B}	406.53 ± 9.3118^{B}	3.6511 ± 0.08905	325.06 ± 6.7858^{b}	2.9392 ± 0.02954^{B}
		p	0.001	0.0007	0.2554	0.0992	0.0023
3: g.15344349A>C	1	AA (67)	9,991.39 ± 192.06	325.46 ± 8.0348^{a}	3.2756 ± 0.07773	299.91 ± 5.8571	3.0135 ± 0.02668^{Aa}
		AC (397)	10005 ± 179.51	326.98 ± 7.5972^{ab}	3.2894 ± 0.07293	297.03 ± 5.5368	2.9797 ± 0.02461^{ABb}
		CC (461)	9,974.55 ± 177.08	322.54 ± 7.512^{b}	3.2564 ± 0.072	295.84 ± 5.4744	2.9781 ± 0.02423^{aBb}
		p	0.8101	0.0497	0.2086	0.248	0.0306
	2	AA (43)	11517 ± 239.39^{A}	421.11 ± 10.051^{A}	3.6242 ± 0.09705	332.59 ± 7.3267^{a}	2.8775 ± 0.03295^{A}
		AC (268)	11148 ± 221.77^{B}	414.03 ± 9.4388^{A}	3.4796 ± 0.09035	326.61 ± 6.8786^{ab}	2.9306 ± 0.02999^{B}
		CC (322)	11050 ± 218.52^{B}	406.27 ± 9.3117^{B}	3.6527 ± 0.08905	324.67 ± 6.7858^{b}	2.9389 ± 0.02954^{B}
		p	0.0004	0.0002	0.3358	0.0539	0.0026
3: g.15345216C>T	1	CC (67)	9,991.39 ± 192.06	325.46 ± 8.0348^{ab}	3.2756 ± 0.07773	299.91 ± 5.8571	3.0135 ± 0.02668^{Aa}
		CT (397)	10005 ± 179.51	326.98 ± 7.5972^{a}	3.2894 ± 0.07293	297.03 ± 5.5368	2.9797 ± 0.02461^{ABb}
		TT (461)	9,974.55 ± 177.08	322.54 ± 7.512^{b}	3.2564 ± 0.072	295.84 ± 5.4744	2.9781 ± 0.02423^{aBb}
		p	0.8101	0.0497	0.2086	0.248	0.0306
	2	CC (43)	11517 ± 239.39^{A}	421.11 ± 10.051^{A}	3.6242 ± 0.09705	332.59 ± 7.3267^{a}	2.8775 ± 0.03295^{A}
		CT (268)	11148 ± 221.77^{B}	414.03 ± 9.4388^{A}	3.6796 ± 0.09035	326.61 ± 6.8786^{ab}	2.9306 ± 0.02999^{B}
		TT (322)	11050 ± 218.52^{B}	406.27 ± 9.3117^{B}	3.6527 ± 0.08905	324.67 ± 6.7858^{b}	2.9389 ± 0.02954^{B}
		p	0.0004	0.0002	0.3358	0.0539	0.0026
3: g.15345227T>C	1	CC (462)	9,976.27 ± 177.07	322.64 ± 7.5119^{a}	3.2568 ± 0.072	295.89 ± 5.4744	2.9781 ± 0.02422^{Aab}
		CT (396)	10001 ± 179.52	326.76 ± 7.5976^{b}	3.2888 ± 0.07294	296.91 ± 5.5371	2.9797 ± 0.02462^{AaB}
		TT (67)	9,990.25 ± 192.06	325.4 ± 8.0349^{ab}	3.2754 ± 0.07773	299.87 ± 5.8572	3.0135 ± 0.02668^{Bb}
		p	0.8681	0.0746	0.2289	0.2745	0.0306
	2	CC (323)	11052 ± 218.52^{A}	406.42 ± 9.3116^{A}	3.6535 ± 0.08905	324.72 ± 6.7857^{a}	2.9388 ± 0.02953^{A}
		CT (267)	11144 ± 221.8^{A}	413.65 ± 9.4399^{B}	3.6778 ± 0.09036	326.5 ± 6.8794^{ab}	2.9307 ± 0.03^{A}
		TT (43)	11516 ± 239.39^{B}	420.99 ± 10.0512^{B}	3.6236 ± 0.09705	332.56 ± 7.3268^{b}	2.8775 ± 0.03295^{B}
		p	0.0005	0.0004	0.3808	0.0596	0.0026
3: g.15349740A>G	1	AA (66)	9,984.83 ± 192.32	324.8 ± 8.0442^{ab}	3.271 ± 0.07783	299.64 ± 5.864	3.0128 ± 0.02672^{a}

(Continued on following page)

TABLE 2 (*Continued*) Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
		AG (398)	10007 ± 179.5	327.15 ± 7.5969^{a}	3.2906 ± 0.07293	297.11 ± 5.5366	2.98 ± 0.02461^{b}
		GG (461)	9,975.25 ± 177.08	322.61 ± 7.5122^{b}	3.2569 ± 0.07201	295.86 ± 5.4746	2.9781 ± 0.02423^{b}
		p	0.7958	0.0436	0.1938	0.2894	0.0383
	2	AA (42)	11509 ± 239.99^{A}	422.11 ± 10.0716^{A}	3.6379 ± 0.09728	332.38 ± 7.3417^{a}	2.8778 ± 0.03305^{A}
		AG (269)	11152 ± 221.74^{B}	413.8 ± 9.4379^{A}	3.6757 ± 0.09034	326.7 ± 6.8779^{ab}	2.9302 ± 0.02999^{B}
		GG (322)	11052 ± 218.52^{B}	406.17 ± 9.3116^{B}	3.651 ± 0.08905	324.71 ± 6.7857^{b}	2.9387 ± 0.02953^{B}
		p	0.0006	0.0001	0.4922	0.0649	0.0031
3: g.15349768A>G	1	AA (66)	9,984.83 ± 192.32	324.8 ± 8.0442^{ab}	3.271 ± 0.07783	299.64 ± 5.864	3.0128 ± 0.02672^{a}
		AG (398)	10007 ± 179.5	327.15 ± 7.5969^{a}	3.2906 ± 0.07293	297.11 ± 5.5366	2.98 ± 0.02461^{b}
		GG (461)	9,975.25 ± 177.08	322.61 ± 7.5122^{b}	3.2569 ± 0.07201	295.86 ± 5.4746	2.9781 ± 0.02423^{b}
		p	0.7958	0.0436	0.1938	0.2894	0.0383
	2	AA (42)	11509 ± 239.99^{A}	422.11 ± 10.0716^{A}	3.6379 ± 0.09728	332.38 ± 7.3417^{a}	2.8778 ± 0.03305^{A}
		AG (269)	11152 ± 221.74^{B}	413.8 ± 9.4379^{A}	3.6757 ± 0.09034	326.7 ± 6.8779^{ab}	2.9302 ± 0.02999^{B}
		GG (322)	11052 ± 218.52^{B}	406.17 ± 9.3116^{B}	3.651 ± 0.08905	324.71 ± 6.7857^{b}	2.9387 ± 0.02953^{B}
		p	0.0006	0.0001	0.4922	0.0649	0.0031
3: g.15349978A>G	1	AA (66)	9,983.68 ± 192.33	324.74 ± 8.0443^{ab}	3.2708 ± 0.07783	299.6 ± 5.8641	3.0128 ± 0.02672^{a}
		AG (397)	10003 ± 179.51	326.94 ± 7.5972^{a}	3.29 ± 0.07293	297 ± 5.5368	2.98 ± 0.02461^{b}
		GG (462)	9,976.97 ± 177.08	322.71 ± 7.5121^{b}	3.2573 ± 0.07201	295.91 ± 5.4745	2.9781 ± 0.02423^{b}
		p	0.8549	0.0661	0.213	0.322	0.0383
	2	AA (42)	11508 ± 239.99^{A}	421.99 ± 10.0718^{A}	3.6373 ± 0.09782	332.34 ± 7.3419^{a}	2.8778 ± 0.03305^{A}
		AG (268)	11148 ± 221.77^{B}	413.42 ± 9.439^{A}	3.6739 ± 0.09035	326.59 ± 6.8787^{ab}	2.9303 ± 0.02999^{B}
		GG (323)	11053 ± 218.51^{B}	406.32 ± 9.3115^{B}	3.6518 ± 0.08905	324.75 ± 6.7856^{b}	2.9387 ± 0.02953^{B}
		p	0.0007	0.0003	0.5521	0.072	0.0032
3: g.15350548C>T	1	CC (66)	9,984.83 ± 192.32	324.8 ± 8.0442^{ab}	3.271 ± 0.07783	299.64 ± 5.864	3.0128 ± 0.02672^{a}
		CT (398)	10007 ± 179.5	327.15 ± 7.5969^{a}	3.2906 ± 0.07293	297.11 ± 5.5366	2.98 ± 0.02461^{b}
		TT (461)	9,975.25 ± 177.08	322.61 ± 7.5122^{b}	3.2569 ± 0.07201	295.86 ± 5.4746	2.9781 ± 0.02423^{b}
		p	0.7958	0.0436	0.1938	0.2894	0.0383
	2	CC (42)	11509 ± 239.99^{A}	422.11 ± 10.0716^{A}	3.6379 ± 0.09728	332.38 ± 7.3417^{a}	2.8778 ± 0.03305^{A}
		CT (269)	11152 ± 221.74^{B}	413.8 ± 9.4379^{A}	3.6757 ± 0.09034	326.7 ± 6.8779^{ab}	2.9302 ± 0.02999^{B}
		TT (322)	11052 ± 218.52^{B}	406.17 ± 9.3116^{B}	3.651 ± 0.08905	324.71 ± 6.7857^{b}	2.9387 ± 0.02953^{B}
		p	0.0006	0.0001	0.4922	0.0649	0.0031
3: g.15350655A>G	1	AA (66)	9,984.83 ± 192.32	324.8 ± 8.0442^{ab}	3.271 ± 0.07783	299.64 ± 5.864	3.0128 ± 0.02672^{a}
		AG (398)	10007 ± 179.5	327.15 ± 7.5969^{a}	3.2906 ± 0.07293	297.11 ± 5.5366	2.98 ± 0.02461^{b}
		GG (461)	9,975.25 ± 177.08		3.2569 ± 0.07201	295.86 ± 5.4746	2.9781 ± 0.02423^{b}

(Continued on following page)

TABLE 2 (*Continued*) Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
				322.61 ± 7.5122^{b}			
		p	0.7958	0.0436	0.1938	0.2894	0.0383
	2	AA (42)	11509 ± 239.99^{A}	422.11 ± 10.0716^{A}	3.6379 ± 0.09728	332.38 ± 7.3417^{a}	2.8778 ± 0.03305^{A}
		AG (269)	11152 ± 221.74^{B}	413.8 ± 9.4379^{A}	3.6757 ± 0.09034	326.7 ± 6.8779^{ab}	2.9302 ± 0.02999^{B}
		GG (322)	11052 ± 218.52^{B}	406.17 ± 9.3116^{B}	3.651 ± 0.08905	324.71 ± 6.7857^{b}	2.9387 ± 0.02953^{B}
		p	0.0006	0.0001	0.4922	0.0649	0.0031
3: g.15350805T>C	1	CC (461)	9,975.25 ± 177.08	322.61 ± 7.5122^{a}	3.2569 ± 0.07201	295.86 ± 5.4746	2.9781 ± 0.02423^{a}
		CT (398)	10007 ± 179.5	327.15 ± 7.5969^{b}	3.2906 ± 0.07293	297.11 ± 5.5366	2.98 ± 0.02461^{a}
		TT (66)	9,984.83 ± 192.32	324.8 ± 8.0442^{ab}	3.271 ± 0.07783	299.64 ± 5.864	3.0128 ± 0.02672^{b}
		p	0.7958	0.0436	0.1938	0.2894	0.0383
	2	CC (322)	11052 ± 218.52^{A}	406.17 ± 9.3116^{A}	3.651 ± 0.08905	324.71 ± 6.7857^{a}	2.9387 ± 0.02953^{A}
		CT (269)	11152 ± 221.74^{A}	413.8 ± 9.4379^{B}	3.6757 ± 0.09034	326.7 ± 6.8779^{ab}	2.9302 ± 0.02999^{A}
		TT (42)	11509 ± 239.99^{B}	422.11 ± 10.0716^{B}	3.6379 ± 0.09728	332.38 ± 7.3417^{b}	2.8778 ± 0.03305^{B}
		p	0.0006	0.0001	0.4922	0.0649	0.0031
3: g.15350898T>C	1	CC (26)	9,605.25 ± 218.94^{Aab}	310.38 ± 8.9936^{Aa}	3.2436 ± 0.08808	284.92 ± 6.5587^{A}	2.9832 ± 0.03096^{ab}
		CT (286)	10067 ± 179.53^{aB}	326.76 ± 7.5947^{aBb}	3.2677 ± 0.07293	297.89 ± 5.535^{B}	2.9717 ± 0.02464^{a}
		TT (613)	9,958.02 ± 177.04^{Bb}	323.02 ± 7.5115^{ABb}	3.2673 ± 0.07199	296.45 ± 5.4741^{B}	2.9886 ± 0.02421^{b}
		p	0.0016	0.0042	0.9024	0.0035	0.0981
	2	CC (18)	11103 ± 279.57	414.38 ± 11.5034	3.7121 ± 0.1126	328.79 ± 8.3889	2.9547 ± 0.03933^{ab}
		CT (189)	11144 ± 220.3	407.97 ± 9.371	3.6433 ± 0.08972	325.12 ± 6.8292	2.917 ± 0.02985^{a}
		TT (426)	11126 ± 219.59	411.38 ± 9.3557	3.6621 ± 0.08949	326.94 ± 6.8179	2.9373 ± 0.02967^{b}
		p	0.9493	0.34	0.5521	0.548	0.0781
3: g.15352855T>C	1	CC (461)	9,970.2 ± 177.08	322.5 ± 7.512^{A}	3.2577 ± 0.072	295.73 ± 5.4745	2.9783 ± 0.02423^{a}
		CT (398)	10018 ± 179.5	327.41 ± 7.5968^{B}	3.2889 ± 0.07293	297.41 ± 5.5365	2.9795 ± 0.02461^{a}
		TT (66)	9,988.28 ± 192.32	324.87 ± 8.044^{AB}	3.2704 ± 0.07783	299.73 ± 5.8639	3.0126 ± 0.02672^{b}
		p	0.5913	0.0258	0.2449	0.2005	0.0391
	2	CC (321)	11060 ± 218.51^{A}	406.48 ± 9.3115^{A}	3.6516 ± 0.08905	325.01 ± 6.7856^{a}	2.9396 ± 0.02953^{A}
		CT (270)	11131 ± 221.71^{A}	412.95 ± 9.4368^{B}	3.6742 ± 0.09032	325.91 ± 6.8771^{ab}	2.9285 ± 0.02998^{A}
		TT (42)	11503 ± 239.98^{B}	421.83 ± 10.0714^{B}	3.6374 ± 0.09728	332.12 ± 7.3416^{b}	2.8773 ± 0.03305^{B}
		p	0.0012	0.0006	0.5375	0.1129	0.0023
3: g.15353088A>C	1	AA (66)	9,986.16 ± 192.32	324.85 ± 8.0442^{AB}	3.271 ± 0.07783	299.67 ± 5.864	3.0127 ± 0.02672^{a}
		AC (399)	10011 ± 179.49	327.31 ± 7.5967^{A}	3.2906 ± 0.07293	297.22 ± 5.5365	2.9798 ± 0.02461^{b}
		CC (460)	9,973.32 ± 177.08	322.54 ± 7.5122^{B}	3.2569 ± 0.07201	295.81 ± 5.4746	2.9782 ± 0.02423^{b}
		p	0.7219	0.0317	0.1945	0.2565	0.0387

(Continued on following page)

TABLE 2 (*Continued*) Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
	2	AA (42)	11503 ± 239.98^{A}	421.83 ± 10.0714^{A}	3.6374 ± 0.09728	332.12 ± 7.3416^{a}	2.8773 ± 0.03305^{A}
		AC (270)	11131 ± 221.71^{B}	412.95 ± 9.4368^{A}	3.6742 ± 0.09032	325.91 ± 6.8771^{ab}	2.9285 ± 0.02998^{B}
		CC (321)	11060 ± 218.51^{B}	406.48 ± 9.3115^{B}	3.6516 ± 0.08905	325.01 ± 6.7856^{b}	2.9396 ± 0.02953^{B}
		p	0.0012	0.0006	0.5375	0.1129	0.0023
3: g.15353235T>C	1	CC (458)	9,975.3 ± 177.07	322.62 ± 7.5117^{A}	3.2571 ± 0.072	295.91 ± 5.4742	2.9786 ± 0.02422^{Aab}
		CT (400)	10020 ± 179.44	327.58 ± 7.5948^{B}	3.2902 ± 0.07291	297.41 ± 5.535	2.9789 ± 0.02461^{AaB}
		TT (67)	9,958.69 ± 192.25	323.73 ± 8.042^{AB}	3.2698 ± 0.0778	298.9 ± 5.8624	3.0139 ± 0.02671^{Bb}
		p	0.5843	0.0224	0.2057	0.3534	0.0279
	2	CC(320)	110073 ± 218.46^{A}	406.61 ± 9.309^{A}	3.6482 ± 0.08903	325.26 ± 6.7838	2.9382 ± 0.02953^{A}
		CT (270)	11132 ± 221.57^{A}	413.3 ± 9.4313^{B}	3.6768 ± 0.09027	325.94 ± 6.8731	2.9284 ± 0.02996^{B}
		TT (43)	11463 ± 239.82^{B}	421.26 ± 10.0661^{B}	3.6472 ± 0.09722	331.25 ± 7.3377	2.8812 ± 0.03302^{B}
		p	0.0046	0.0006	0.4487	0.2026	0.0054
3: g.15353254T>C	1	CC (458)	9,975.3 ± 177.07	322.62 ± 7.5117^{A}	3.2571 ± 0.072	295.91 ± 5.4742	2.9786 ± 0.02422^{Aab}
		CT (400)	10020 ± 179.44	327.58 ± 7.5948^{B}	3.2902 ± 0.07291	297.41 ± 5.535	2.9789 ± 0.02461^{AaB}
		TT (67)	9,958.69 ± 192.25	323.73 ± 8.042^{AB}	3.2698 ± 0.0778	298.9 ± 5.8624	3.0139 ± 0.02671^{Bb}
		p	0.5843	0.0224	0.2057	0.3534	0.0279
	2	CC (320)	11073 ± 218.46^{A}	406.61 ± 9.309^{A}	3.6482 ± 0.08903	325.26 ± 6.7838	2.9382 ± 0.02953^{A}
		CT (270)	11132 ± 221.57^{A}	413.3 ± 9.4313^{B}	3.6768 ± 0.09027	325.94 ± 6.8731	2.9284 ± 0.02996^{A}
		TT (43)	11463 ± 239.82^{B}	421.26 ± 10.0661^{B}	3.6472 ± 0.09722	331.25 ± 7.3377	2.8812 ± 0.03302^{B}
		p	0.0046	0.0006	0.4487	0.2026	0.0054
3: g.15353292C>G	1	CC (70)	9,944.8 ± 191.52	322.37 ± 8.016^{AB}	3.2614 ± 0.07752	298.06 ± 5.8434	3.01 ± 0.02659^{a}
		CG (398)	10030 ± 179.48	328.03 ± 7.5959^{A}	3.2913 ± 0.07292	297.73 ± 5.5359	2.9791 ± 0.02461^{b}
		GG (457)	9,974.5 ± 177.06	322.75 ± 7.5113^{B}	3.2585 ± 0.072	295.96 ± 5.4739	2.9792 ± 0.02422^{b}
		p	0.4024	0.0097	0.1968	0.3649	0.0565
	2	CC (45)	11416 ± 238.71^{Aa}	419.51 ± 10.0282^{A}	3.6449 ± 0.0968	330.41 ± 7.31	2.8865 ± 0.03283^{Aa}
		CG (268)	11140 ± 221.63^{ABb}	413.68 ± 9.4334^{A}	3.6778 ± 0.09029	326.11 ± 6.8746	2.9276 ± 0.02997^{ABb}
		GG (320)	11077 ± 218.46^{aBb}	406.77 ± 9.3091^{B}	3.6485 ± 0.08903	325.33 ± 6.7838	2.9379 ± 0.02953^{aBb}
		p	0.0135	0.0012	0.4243	0.302	0.0114
3: g.15353330A>G	1	AA (79)	9,899.56 ± 189.6	322.46 ± 7.9485^{AB}	3.2788 ± 0.07679	297 ± 5.794	3.013 ± 0.02628^{A}
		AG (389)	10046 ± 179.54	328.15 ± 7.5982^{A}	3.2868 ± 0.07295	298.05 ± 5.5376	2.9774 ± 0.02462^{B}
		GG (457)	9,979.48 ± 177.05	322.76 ± 7.511^{B}	3.2568 ± 0.07199	296.07 ± 5.4737	2.9787 ± 0.02422^{B}
		p	0.1389	0.0081	0.276	0.3331	0.0153
	2	AA (50)	11426 ± 236.22^{A}	421.24 ± 9.9409^{A}	3.6591 ± 0.09585	331.51 ± 7.2461^{a}	2.8942 ± 0.03242^{Aa}

(Continued on following page)

TABLE 2 (*Continued*) Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
		AG (263)	11129 ± 221.8^{B}	412.97 ± 9.4393^{A}	3.6745 ± 0.09036	325.64 ± 6.879^{ab}	2.9267 ± 0.03^{ABb}
		GG (320)	11074 ± 218.47^{B}	406.52 ± 9.3093^{B}	3.6473 ± 0.08903	325.18 ± 6.784^{b}	2.9375 ± 0.02953^{aBb}
		p	0.0063	0.0005	0.5163	0.1237	0.0275
3: g.15353342C>T	1	CC (82)	9,920.86 ± 189.11	321.48 ± 7.9312^{AaB}	3.2598 ± 0.0766	297.12 ± 5.7813	3.0075 ± 0.0262^{a}
		CT (387)	10045 ± 179.6	328.74 ± 7.6001^{Ab}	3.2934 ± 0.07297	298.09 ± 5.5389	2.9783 ± 0.02463^{b}
		TT (456)	9,975.84 ± 177.05	322.8 ± 7.5109^{aBb}	3.2587 ± 0.07199	296.03 ± 5.4736	2.9795 ± 0.02422^{b}
		p	0.1847	0.002	0.1603	0.3057	0.0546
	2	CC (50)	11394 ± 236.17^{Aa}	417.79 ± 9.9384^{AaB}	3.6365 ± 0.09582	330.51 ± 7.2443	2.8945 ± 0.03241^{A}
		CT (263)	11138 ± 221.8^{ABb}	414.03 ± 9.4395^{Aab}	3.6814 ± 0.09036	325.94 ± 6.8791	2.9266 ± 0.03^{AB}
		TT (320)	11077 ± 218.46^{aBb}	406.92 ± 9.3091^{Bb}	3.6496 ± 0.08903	325.3 ± 6.7838	2.9374 ± 0.02953^{B}
		p	0.0157	0.0019	0.3239	0.2462	0.0285
3: g.15355389T>C	1	CC (455)	9,959.33 ± 176.95^{ab}	321.99 ± 7.5076^{Aab}	3.257 ± 0.07196	295.75 ± 5.4712^{ab}	2.9821 ± 0.0242
		CT (358)	10028 ± 179.57^{a}	327.84 ± 7.5987^{aB}	3.2897 ± 0.07295	298.36 ± 5.5379^{a}	2.9864 ± 0.02463
		TT (109)	9,872.62 ± 186.83^{b}	321.17 ± 7.854^{ABb}	3.2816 ± 0.07574	293.97 ± 5.7248^{b}	2.9925 ± 0.02581
		p	<0.0001	<0.0001	0.3159	0.0002	0.0374
	2	CC (317)	11088 ± 218.46	408.02 ± 9.3087^{A}	3.656 ± 0.08903^{ab}	325.66 ± 6.7835	2.9377 ± 0.02953^{A}
		CT (241)	11207 ± 222.08	417.82 ± 9.4492^{B}	3.6914 ± 0.09046^{a}	327.92 ± 6.8862	2.926 ± 0.03005^{AB}
		TT (71)	11213 ± 231.31	407.7 ± 9.7655^{A}	3.6031 ± 0.09396^{b}	325.7 ± 7.1178	2.8987 ± 0.03161^{B}
		p	0.2155	0.0003	0.1161	0.4657	0.0585
3: g.15355514T>C	1	CC (457)	9,985.46 ± 176.88^{ab}	322.9 ± 7.5049^{A}	3.2557 ± 0.07193	296.38 ± 5.4692^{ab}	2.9799 ± 0.0242
		CT (358)	10045 ± 179.49^{a}	328.34 ± 7.5957^{B}	3.2879 ± 0.07292	298.76 ± 5.5357^{a}	2.9847 ± 0.02462
		TT (108)	9,866.46 ± 187.14^{b}	320.88 ± 7.8657^{A}	3.2799 ± 0.07586	293.88 ± 5.7333^{b}	2.9928 ± 0.02586
		p	0.0004	0.0003	0.3029	0.0044	0.0152
	2	CC (318)	11079 ± 218.46^{a}	407.65 ± 9.3088^{Aab}	3.6551 ± 0.08903^{ab}	325.34 ± 6.7836^{a}	2.9368 ± 0.02953^{a}
		CT (240)	11229 ± 222.11^{b}	418.7 ± 9.4505^{aB}	3.6928 ± 0.09048^{a}	328.76 ± 6.8872^{b}	2.9277 ± 0.03005^{ab}
		TT (72)	11227 ± 231.13^{ab}	409.02 ± 9.7594^{ABb}	3.6098 ± 0.09389^{b}	326.41 ± 7.1133^{ab}	2.9012 ± 0.03158^{b}
		p	0.0866	<0.0001	0.0542	0.2667	0.0883
3: g.15355833A>G	1	AA (107)	9,845.22 ± 187.08^{a}	320.15 ± 7.8632^{AaB}	3.2817 ± 0.07583	293.31 ± 5.7315^{a}	2.9945 ± 0.02585
		AG (358)	10007 ± 179.62^{b}	327.02 ± 7.601^{Ab}	3.2898 ± 0.07298	297.7 ± 5.5396^{b}	2.9868 ± 0.02464
		GG (458)	9,951 ± 176.98^{ab}	321.72 ± 7.509^{aBb}	3.2577 ± 0.07197	295.45 ± 5.4722^{ab}	2.9822 ± 0.02421
		p	<0.0001	<0.0001	0.3381	<0.0001	0.0743
	2	AA (71)	11216 ± 231.27^{ab}		3.6059 ± 0.09395^{a}	325.9 ± 7.117	2.8997 ± 0.03161^{a}

(Continued on following page)

TABLE 2 (*Continued*) Associations of 21 SNPs in *PKLR* with milk production traits in two lactations of Chinese Holstein cows (LSM ±SE).

SNP name	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
				408.03 ± 9.7644^{A}			
		AG (240)	11229 ± 222.11^{a}	418.77 ± 9.4505^{B}	3.6936 ± 0.09048^{b}	328.77 ± 6.8871	2.9279 ± 0.03005^{ab}
		GG (319)	11082 ± 218.45^{b}	407.55 ± 9.3085^{A}	3.6525 ± 0.08902^{ab}	325.37 ± 6.7834	2.936 ± 0.02953^{b}
		p	0.1047	<0.0001	0.0553	0.1983	0.0436

Note: LSM ±SE: Least Squares Mean ± Standard Deviation; the number in the bracket represents the number of cows for the corresponding genotype; *p* shows the significance for the genetic effects of SNPs; a, b within the same column with different superscripts means $p < 0.05$; and A, B within the same column with different superscripts means $p < 0.01$.

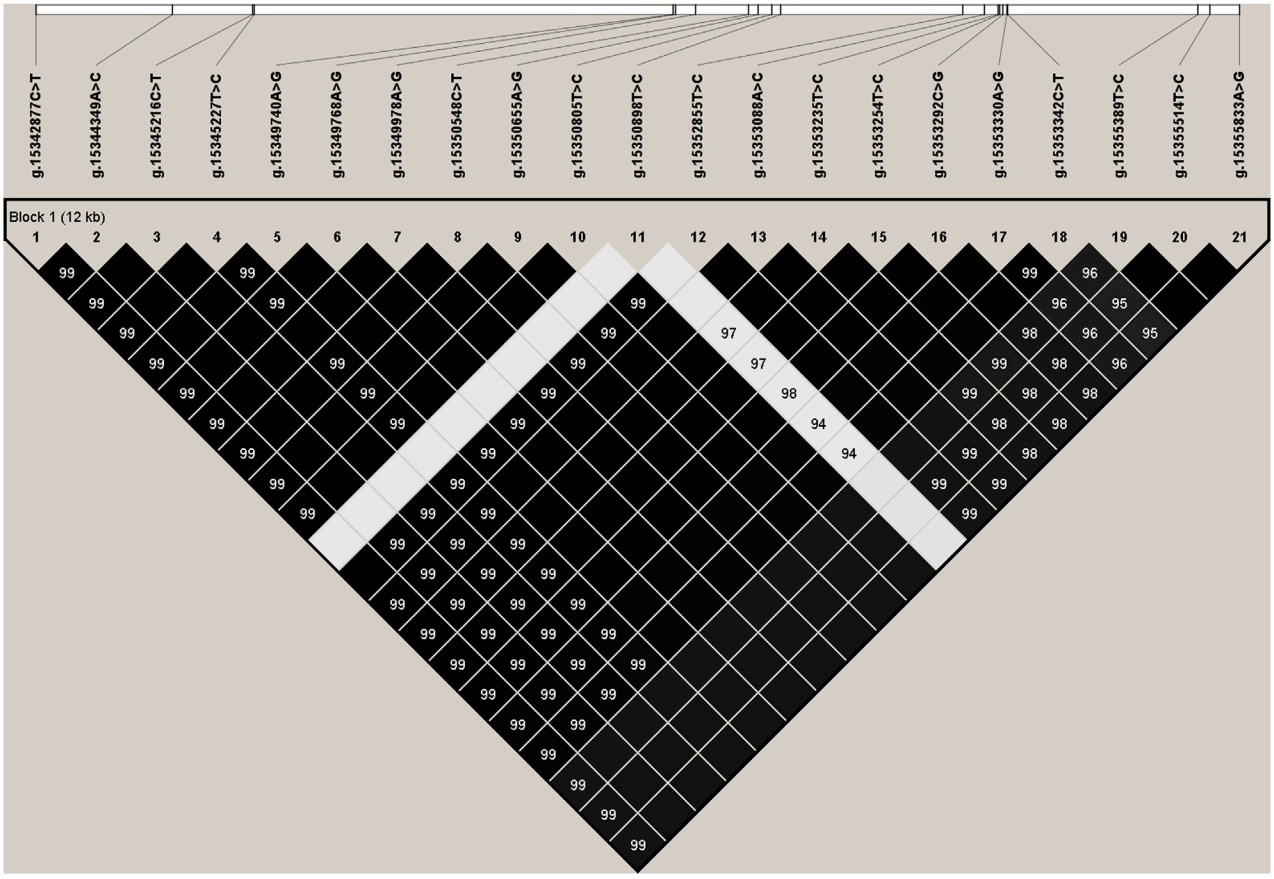


FIGURE 1
Linkage disequilibrium estimated between SNPs in *PKLR* gene. The values in the black boxes are pair-wise SNP correlations (R^2).

increase the MFE to cause the instability of *PKLR* mRNA secondary structure, and the other seven sites, 3:g.15349978G, 3:g.15350655G, 3:g.15352855C, 3:g.15353088C, 3:g.15353292G, 3:g.15353330G, and 3:g.15353342T, could decrease the MFE and make the mRNA secondary structure more stable.

Discussion

Our previous study considered *PKLR* gene to be a candidate to affect milk production traits in dairy cattle (Liang et al., 2017). In this study, we identified totally 21 SNPs in *PKLR* gene, and

TABLE 3 Changes in transcription factors binding sites (TFBSs) caused by the SNPs in 5′ regulatory region of *PKLR*.

SNP name	Allele	TFs	*p*	Predicted core binding site sequence
3:g.15342877C>T	C	SP100	0.0030	TCCGTCGCTTAAAAG
		ESRRA	0.0046	TAGGTCAGTCAAGGTCA
	T	—	—	—
3:g.15344349A>C	A	MLX	0.0034	ATCACGTGAT
		ZBTB33	0.0042	CTCTCGCGAGATCTG
		IRF5	0.0048	TTGATCGAGAATTCC
	C	ZNF524	0.0014	ACCCTCGAACCC
		YY2	0.0021	CCATGCCGCCAT
		SREBF2	0.0044	ATCACGTGAC
3:g.15345216C>T	C	RREB1	0.0031	CCCCAAACCACCCCCCCCC
	T	TWIST2	0.0008	CGCAGCTGCG
		ZEB1	0.0009	CCCACCTGCGC
		NAC007	0.0010	GCCAGCTGGC
		BHLHE22	0.0016	CGCAGCTGCG
		ZFP42	0.0016	GTTCCAAAATGGCTGCCTCCG
		TCF3	0.0024	CGCACCTGCCC
		NAC031	0.0029	AGCAGCTGCT
		ZSCAN31	0.0030	GCATAACTGCCCTGCGTCC
		TCF12	0.0046	CGCACCTGCCG
3:g.15345227T>C	T	—	—	—
	C	MYC	0.0029	GGCCACGTGCCC
		TFAP2A	0.0031	ATTGCCTCAGGCCA
		TCF4	0.0032	CGGCACCTGCCCC

Note: TFs: transcription factors; SNP, site is underlined.

TABLE 4 The minimum free energy (MFE) values of optimal secondary structure of *PKLR* mRNA.

Mutant site	MFE (kcal/mol)
References sequence	−1,145.2
3:g.15345216T	−1,143
3:g.15345227C	−1,144.9
3:g.15349768G	−1,144.5
3:g.15349978G	−1,146
3:g.15350655G	−1,148.80
3:g.15350898C	−1,143.3
3:g.15352855C	−1,145.6
3:g.15353088C	−1,147.20
3:g.15353235C	−1,144.6
3:g.15353254C	−1,143.90
3:g.15353292G	−1,149.3
3:g.15353330G	−1,151.80
3:g.15353342T	−1,150

Note: MFE: minimum free energy; reference sequence: XM_024989616.1 (ARS-UCD1.2).

found that all the SNPs were significantly associated with at least one milk production trait, simultaneously, the results of haplotype association analysis were basically consistent with the single marker association analysis, which suggested that the *PKLR* gene had large genetic effect on milk production traits. Brondum et al. (2015) added the sequence data of a few significant variation into the conventional 54k SNPs for single marker analysis, and found it can improve the reliability of genomic prediction, for instance, the reliability of the Nordic Holstein cattle milk production traits increased by 4%, that of Nordic red bull increased by 3%, and that of France Holstein cows increased by 5%. Currently, four commercial gene chips, including illumina Bovine SNP50K BeadChip, illumina BovineHD Genotyping BeadChip, GeneSeek Genomic Profiler (GGP) Bovine 150K, and 100K arrays, do not contain SNPs identified in this study, after that, we could try to add significant functional SNPs in this study to gene chips to improve the accuracy of genomic prediction in dairy cattle.

PKLR converts phosphoenolpyruvic acid to pyruvate, the main carbon source, and its perturbation may significantly affect

the pyruvate levels in cells (Liu et al., 2019). Moreover, pyruvate is an important intermediate in the glucose metabolism of all living organisms and the mutual transformation of various substances in the body. It can also convert sugars, fats and amino acids into each other through acetyl CoA and the tricarboxylic acid cycle (Gray et al., 2014). Studies have shown that *PKLR* regulates and influences key metabolic pathways related to lipid metabolism, steroid biosynthesis, PPAR signaling pathway, fatty acid synthesis and oxidation (Lee et al., 2017; Mardinoglu et al., 2018; Liu et al., 2019). It can be seen that *PKLR* gene can regulate the synthesis of milk components, especially milk lactose and fat.

Transcription factors are a group of protein molecules that bind to TFBSs to ensure that the target gene is expressed at a specific intensity at a specific time and space (Jolma et al., 2013). When the mutation site changes that it will affect the binding of TFs to TFBSs, and then inhibiting or enhancing gene expression (Spivakov et al., 2012). In this study, four SNPs in 5′ region of *PKLR* were predicted to change the TFBSs that would be affect the expression of the downstream gene. For the 3:g.15342877C>T, the allele C could bind SP100 and ESRRA, and the milk and fat yields of CC genotype cows was significantly higher than that of TT individuals. In addition, it has reported that ESRRA enhanced the transcriptional activation of numerous autophagy-related (Atg) genes, *Atg5*, *Atg16l1*, and *Becn1* (Kim et al., 2018). SP100 may function as a nuclear hormone receptor transcriptional coactivator (Bloch et al., 2000). It can be inferred that the increase of CC genotype phenotype may be due to the combination of transcription factors SP100 and ESRRA at the C site, which together activate the expression of gene *PKLR*. The allele A in 3:g.15344349A>C could bind MLX, ZBTB33, and IRF5, and the allele C binds ZNF524, YY2, and SREBF2, meanwhile, the milk and fat yields of AA genotype cows was significantly higher than that of CC individuals. MLX plays a role in transcriptional activation of glycolytic target genes and the Mondo family (Billin et al., 2000; Sans et al., 2006). ZBTB33 activated transcription from exogenous methylated promoters (Zhenilo et al., 2018). IRF5 directly activated transcription of the genes *IL-12p40*, *IL-12p35,* and *IL-23p19* and contributed to the plasticity of macrophage polarization (Krausgruber et al., 2011). YY2 reduces the activity of the *c-Myc* and *CXCR4* promoter (Nguyen et al., 2004). SREBF2 can activate the transcription of genes involved in cholesterol biosynthesis (Xu et al., 2020; Sellers et al., 2021). The functional role of TF ZNF524 is unclear so far. It is speculated that the higher milk yield of AA genotype cows may be the result of combined activation of transcription factors MLX, ZBTB33 and IRF5 or the inhibition of ZNF524, YY2, and SREBF2 on the expression of gene *PKLR*. For the 3:g.15345216C>T, the allele C binds RREB1, and allele T could bind TWIST2, ZEB1, NAC007, BHLHE22, ZFP42, TCF3, NAC031, ZSCAN31, and TCF12, as well as, the milk and fat yields of CC genotype cows was significantly higher than that of TT individuals. RREB1 is a transcriptional activator of calcitonin in response to Ras signaling (Deng et al., 2020). TWIST2 can suppress the expression of *FGF21* to activate the AMPK/mTOR signalling pathway which inhibits the progression of various cancers (Song et al., 2021). ZEB1 as a direct transcriptional repressor of E-cadherin by physically binding to the proximal promoter of E-cadherin in breast cancers (Eger et al., 2005). BHLHE22 is a transcriptional repressor and is involved in cell differentiation in neuron development (Ross et al., 2012; Darmawi et al., 2022), TCF3 combined with HDAC3 down-regulates the expression of miR-101 that is a type of tumor suppressor gene, thereby promoting the proliferation of BL cells and inhibiting their apoptosis (Dong et al., 2021). TCF12 functions as transcriptional repressor of E-cadherin (Lee et al., 2012). The function of some transcription factors, NAC007, ZFP42, NAC031, and ZSCAN31, is still unclear. Therefore, it can be speculated that the increased phenotype of CC genotype individuals may be caused by the activation of *PKLR* gene expression by binding the TF RREB1, or the co-inhibition of *PKLR* gene expression by TFs TWIST2, ZEB1, NAC007, BHLHE22, ZFP42, TCF3, NAC031, ZSCAN31, and TCF12. For the 3:g.15345227T>C, the allele C could bind MYC, TFAP2A and TCF4, and the milk and fat yields of CC genotype cows was significantly lower than that of TT individuals. MYC represses transcription when tethered to promoters by Miz1 or other proteins (Adhikary and Eilers 2005). TFAP2A appeared to strengthen the binding of Smad2/3 to target promoters and affect transcriptional responses in knockdown experiments (Koinuma et al., 2009). TCF4 is involved in the initiation of neuronal differentiation by binding to the E box to activate transcription (Teixeira et al., 2021). It can be speculated that the decrease of CC genotype phenotype may be due to the combination of TFs MYC, TFAP2A, and TCF4 to inhibit the expression of *PKLR* gene. Thus, we speculated that these four SNP mutations changed the TFBSs to modulate the gene expression of *PKLR*, resulting in changes of phenotypic data.

The secondary structure of mRNA is formed by the complementary pairing of bases on the single chain, and the same mRNA molecules can be folded to form a variety of configurations. The secondary structure of mRNA, as the skeleton of the higher functions of RNA, plays an important role in various life processes, including protein folding and transport, initiation and extension of translation process, regulation of translation rate and direct influence the stability of mRNA itself (Wan et al., 2011; Dethoff et al., 2012). The base change of SNP may change the secondary structure of mRNA, so we used RNAfold to predict the secondary structure of mRNA, and MFE was used as an indicator to measure the stability of the secondary structure in this study. Five sites, 3:g.15345216T, 3:g.15349768G, 3:g.15350898C, 3:g.15353235C, and 3:g.15353254C, with higher MFEs compared that to the reference sites, caused the instability of *PKLR* mRNA secondary structure to inhibit its expression, additionally, our

study found that the five loci were significantly associated with milk fat yield, and the phenotypic value of fat yield of homozygous individuals at the mutation site was significantly reduced. On the contrary, three sites, 3:g.15353292G, 3:g.15353330G, and 3:g.15353342T, had lower MFEs and more stable mRNA structure, also had significant genetic effects on fat yield, and the phenotypes of fat yield of homozygous cows at these sites were significant increased. It suggested that these eight SNP sites might affect milk fat yield of dairy cows by influencing the instability of mRNA secondary structure of *PKLR*. Further, we speculated that the changes of mRNA secondary structures caused by SNPs may affect the stability of its higher-order structure and gene expression, leading to an influence on milk production phenotypes of dairy cows.

Conclusion

In summary, a total of 21 SNPs were identified in *PKLR* gene, and their significant genetic associations with milk production traits of dairy cows have been confirmed. Eleven SNPs might be the potential causal mutations for the milk production traits in dairy cattle that needs more in-depth validation, of which, 3:g.15342877C>T, 3:g.15344349A>C, 3:g.15345216C>T, and 3:g.15345227T>C might change the TFBSs to regulate expression of the *PKLR* gene, and eight SNPs, 3:g.15345216C>T, 3:g.15349768A>G, 3:g.15350898T>C, 3:g.15353235T>C, 3:g.15353254T>C, 3:g.15353292C>G, 3:g.15353330A>G, and 3:g.15353342C>T, could change the secondary structure of mRNA and the phenotypic value of fat yield. The valuable SNPs could be used as candidate genetic markers for dairy cattle molecular breeding for the development of GS chip.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

BH, DS, and KC: conceptualization, methodology, and funding acquisition. LX and YL: formal analysis. FZ: investigation and resources. AD: visualization. AD: writing—original draft preparation. BH: writing, review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by Shandong Provincial Natural Science Foundation (ZR2020MC165), National Natural Science Foundation of China (32072716, 31872330), China Agriculture Research System of MOF and MARA (CARS-36), and the Program for Changjiang Scholar and Innovation Research Team in University (IRT_15R62).

Acknowledgments

We appreciate Beijing Dairy Cattle Center for providing the semen and blood samples and phenotypic data.

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.2022.1002706/full#supplementary-material

References

Abargouei, A. S., Janghorbani, M., Salehi-Marzijarani, M., and Esmaillzadeh, A. (2012). Effect of dairy consumption on weight and body composition in adults: a systematic review and meta-analysis of randomized controlled clinical trials. *Int. J. Obes.* 36, 1485–1493. doi:10.1038/ijo.2011.269

Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. *Nat. Rev. Mol. Cell Biol.* 6, 635–645. doi:10.1038/nrm1703

Ahrens, M., Ammerpohl, O., von Schonfels, W., Kolarova, J., Bens, S., Itzel, T., et al. (2013). DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* 18, 296–302. doi:10.1016/j.cmet.2013.07.004

Billin, A. N., Eilers, A. L., Coulter, K. L., Logan, J. S., and Ayer, D. E. (2000). MondoA, a novel basic helix-loop-helix-leucine zipper transcriptional activator that constitutes a positive branch of a max-like network. *Mol. Cell. Biol.* 20, 8845–8854. doi:10.1128/mcb.20.23.8845-8854.2000

Bloch, D. B., Nakajima, A., Gulick, T., Chiche, J. D., Orth, D., de La Monte, S. M., et al. (2000). Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. *Mol. Cell. Biol.* 20, 6138–6146. doi:10.1128/mcb.20.16.6138-6146.2000

Brondum, R. F., Su, G., Janss, L., Sahana, G., Guldbrandtsen, B., Boichard, D., et al. (2015). Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98, 4107–4116. doi:10.3168/jds.2014-9005

Chella Krishnan, K., Floyd, R. R., Sabir, S., Jayasekera, D. W., Leon-Mimila, P. V., Jones, A. E., et al. (2021). Liver pyruvate kinase promotes NAFLD/NASH in both mice and humans in a sex-specific manner. *Cell. Mol. Gastroenterol. Hepatol.* 11, 389–406. doi:10.1016/j.jcmgh.2020.09.004

Crichton, G. E., Bryan, J., Buckley, J., and Murphy, K. J. (2011). Dairy consumption and metabolic syndrome: a systematic review of findings and methodological issues. *Obes. Rev.* 12, e190–201. doi:10.1111/j.1467-789X.2010.00837.x

DarmawiChen, L. Y., Su, P. H., Liew, P. L., Wang, H. C., Weng, Y. C., Huang, R. L., et al. (2022). BHLHE22 expression is associated with a proinflammatory immune microenvironment and confers a favorable prognosis in endometrial cancer. *Int. J. Mol. Sci.* 23, 7158. doi:10.3390/ijms23137158

de Las Heras-Saldana, S., Lopez, B. I., Moghaddar, N., Park, W., Park, J. E., Chung, K. Y., et al. (2020). Use of gene expression and whole-genome sequence information to improve the accuracy of genomic prediction for carcass traits in Hanwoo cattle. *Genet. Sel. Evol.* 52, 54. doi:10.1186/s12711-020-00574-2

Deng, Y. N., Xia, Z., Zhang, P., Ejaz, S., and Liang, S. (2020). Transcription factor RREB1: From target genes towards biological functions. *Int. J. Biol. Sci.* 16, 1463–1473. doi:10.7150/ijbs.40834

Dethoff, E. A., Chugh, J., Mustoe, A. M., and Al-Hashimi, H. M. (2012). Functional complexity and regulation through RNA dynamics. *Nature* 482, 322–330. doi:10.1038/nature10885

Dong, L., Huang, J., Zu, P., Liu, J., Gao, X., Du, J., et al. (2021). Transcription factor 3 (TCF3) combined with histone deacetylase 3 (HDAC3) down-regulates microRNA-101 to promote Burkitt lymphoma cell proliferation and inhibit apoptosis. *Bioengineered* 12, 7995–8005. doi:10.1080/21655979.2021.1977557

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., et al. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene* 24, 2375–2385. doi:10.1038/sj.onc.1208429

Fu, Y., Jia, R., Xu, L., Su, D., Li, Y., Liu, L., et al. (2022). Fatty acid desaturase 2 affects the milk-production traits in Chinese Holsteins. *Anim. Genet.* 53, 422–426. doi:10.1111/age.13192

Gebreyesus, G., Buitenhuis, A. J., Poulsen, N. A., Visker, M., Zhang, Q., van Valenberg, H. J. F., et al. (2019). Multi-population GWAS and enrichment analyses reveal novel genomic regions and promising candidate genes underlying bovine milk fatty acid composition. *BMC Genomics* 20, 178. doi:10.1186/s12864-019-5573-9

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. doi:10.1111/j.1439-0388.2007.00702.x

Gray, L. R., Tompkins, S. C., and Taylor, E. B. (2014). Regulation of pyruvate metabolism and human disease. *Cell. Mol. Life Sci.* 71, 2577–2604. doi:10.1007/s00018-013-1539-2

Han, B., Liang, W., Liu, L., Li, Y., and Sun, D. (2017). Determination of genetic effects of ATF3 and CDKN1A genes on milk yield and compositions in Chinese Holstein population. *BMC Genet.* 18, 47. doi:10.1186/s12863-017-0516-4

Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi:10.2135/cropsci2008.08.0512

Jiang, L., Jiang, J., Yang, J., Liu, X., Wang, J., Wang, H., et al. (2013). Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. *BMC Genomics* 14, 131. doi:10.1186/1471-2164-14-131

Jiang, J., Gao, Y., Hou, Y., Li, W., Zhang, S., Zhang, Q., et al. (2016). Whole-genome resequencing of Holstein bulls for indel discovery and identification of genes associated with milk composition traits in dairy cattle. *PLoS One* 11, e0168946. doi:10.1371/journal.pone.0168946

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339. doi:10.1016/j.cell.2012.12.009

Kim, S. Y., Yang, C. S., Lee, H. M., Kim, J. K., Kim, Y. S., Kim, Y. R., et al. (2018). ESRRA (estrogen-related receptor alpha) is a key coordinator of transcriptional and post-translational activation of autophagy to promote innate host defense. *Autophagy* 14, 152–168. doi:10.1080/15548627.2017.1339001

Koinuma, D., Tsutsumi, S., Kamimura, N., Taniguchi, H., Miyazawa, K., Sunamura, M., et al. (2009). Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol. Cell. Biol.* 29, 172–186. doi:10.1128/MCB.01038-08

Korkuc, P., Arends, D., May, K., Konig, S., and Brockmann, G. A. (2021). Genomic loci affecting milk production in German black pied cattle (DSN). *Front. Genet.* 12, 640039. doi:10.3389/fgene.2021.640039

Krausgruber, T., Blazek, K., Smallie, T., Alzabin, S., Lockstone, H., Sahgal, N., et al. (2011). IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. *Nat. Immunol.* 12, 231–238. doi:10.1038/ni.1990

Lee, C. C., Chen, W. S., Chen, C. C., Chen, L. L., Lin, Y. S., Fan, C. S., et al. (2012). TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. *J. Biol. Chem.* 287, 2798–2809. doi:10.1074/jbc.M111.258947

Lee, S., Zhang, C., Liu, Z., Klevstig, M., Mukhopadhyay, B., Bergentall, M., et al. (2017). Network analyses identify liver-specific targets for treating liver diseases. *Mol. Syst. Biol.* 13, 938. doi:10.15252/msb.20177703

Liang, R., Han, B., Li, Q., Yuan, Y., Li, J., and Sun, D. (2017). Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Sci. Rep.* 7, 6396. doi:10.1038/s41598-017-06634-w

Liu, Z., Zhang, C., Lee, S., Kim, W., Klevstig, M., Harzandi, A. M., et al. (2019). Pyruvate kinase L/R is a regulator of lipid metabolism and mitochondrial function. *Metab. Eng.* 52, 263–272. doi:10.1016/j.ymben.2019.01.001

Liu, L., Zhou, J., Chen, C. J., Zhang, J., Wen, W., Tian, J., et al. (2020). GWAS-based identification of new loci for milk yield, fat, and protein in Holstein cattle. *Animals.* 10, 2048. doi:10.3390/ani10112048

Lopdell, T. J., Tiplady, K., Couldrey, C., Johnson, T. J. J., Keehan, M., Davis, S. R., et al. (2019). Multiple QTL underlie milk phenotypes at the CSF2RB locus. *Genet. Sel. Evol.* 51, 3. doi:10.1186/s12711-019-0446-x

Mardinoglu, A., Uhlen, M., and Boren, J. (2018). Broad views of non-alcoholic fatty liver disease. *Cell Syst.* 6, 7–9. doi:10.1016/j.cels.2018.01.004

Maryam, J., Babar, M. E., Nadeem, A., Yaqub, T., and Hashmi, A. S. (2015). Identification of functional consequence of a novel selection signature in CYP11b1 gene for milk fat content in Bubalus bubalis. *Meta Gene* 6, 85–90. doi:10.1016/j.mgene.2015.09.002

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Nayeri, S., Sargolzaei, M., Abo-Ismail, M. K., May, N., Miller, S. P., Schenkel, F., et al. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* 17, 75. doi:10.1186/s12863-016-0386-1

Nguyen, N., Zhang, X., Olashaw, N., and Seto, E. (2004). Molecular cloning and functional characterization of the transcription factor YY2. *J. Biol. Chem.* 279, 25927–25934. doi:10.1074/jbc.M402525200

Ross, S. E., McCord, A. E., Jung, C., Atan, D., Mok, S. I., Hemberg, M., et al. (2012). Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron* 73, 292–303. doi:10.1016/j.neuron.2011.09.035

Rumbold, P., McCullogh, N., Boldon, R., Haskell-Ramsay, C., James, L., Stevenson, E., et al. (2021). The potential nutrition-physical- and health-related benefits of cow's milk for primary-school-aged children. *Nutr. Res. Rev.* 35, 50–69. doi:10.1017/s095442242100007x

Sans, C. L., Satterwhite, D. J., Stoltzman, C. A., Breen, K. T., and Ayer, D. E. (2006). MondoA-mlx heterodimers are candidate sensors of cellular energy status: mitochondrial localization and direct regulation of glycolysis. *Mol. Cell. Biol.* 26, 4863–4871. doi:10.1128/MCB.00657-05

Schrooten, C., Bovenhuis, H., Coppieters, W., and Van Arendonk, J. A. M. (2000). Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci.* 83, 795–806. doi:10.3168/jds.S0022-0302(00)74942-3

Sellers, J., Brooks, A., Fernando, S., Westenberger, G., Junkins, S., Smith, S., et al. (2021). Fasting-Induced upregulation of MKP-1 modulates the hepatic response to feeding. *Nutrients* 13, 3941. doi:10.3390/nu13113941

Shi, L., Liu, L., Lv, X., Ma, Z., Yang, Y., Li, Y., et al. (2019). Polymorphisms and genetic effects of PRLR, MOGAT1, MINPP1 and CHUK genes on milk fatty acid traits in Chinese Holstein. *BMC Genet.* 20, 69. doi:10.1186/s12863-019-0769-1

Soedamah-Muthu, S. S., and de Goede, J. (2018). Dairy consumption and cardiometabolic diseases: systematic review and updated meta-analyses of prospective cohort studies. *Curr. Nutr. Rep.* 7, 171–182. doi:10.1007/s13668-018-0253-y

Song, Y., Zhang, W., Zhang, J., You, Z., Hu, T., Shao, G., et al. (2021). TWIST2 inhibits EMT and induces oxidative stress in lung cancer cells by regulating the FGF21-mediated AMPK/mTOR pathway. *Exp. Cell Res.* 405, 112661. doi:10.1016/j.yexcr.2021.112661

Spivakov, M., Akhtar, J., Kheradpour, P., Beal, K., Girardot, C., Koscielny, G., et al. (2012). Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biol.* 13, R49. doi:10.1186/gb-2012-13-9-r49

Stock, K. F., and Reents, R. (2013). Genomic selection: Status in different species and challenges for breeding. *Reprod. Domest. Anim.* 48, 2–10. doi:10.1111/rda.12201

Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R., and Papes, F. (2021). Transcription factor 4 and its association with psychiatric disorders. *Transl. Psychiatry* 11, 19. doi:10.1038/s41398-020-01138-0

Valsalan, J., Sadan, T., Venkatachalapathy, T., Anilkumar, K., and Aravindakshan, T. V. (2021). Identification of novel single-nucleotide polymorphism at exon1 and 2 region of B4GALT1 gene and its association with milk production traits in crossbred cattle of Kerala, India. *Anim. Biotechnol.* 10, 1–9. doi:10.1080/10495398.2020.1866591

Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., and Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655. doi:10.1038/nrg3049

Wang, H., Maechler, P., Antinozzi, P. A., Hagenfeldt, K. A., and Wollheim, C. B. (2000). Hepatocyte nuclear factor 4alpha regulates the expression of pancreatic beta -cell genes implicated in glucose metabolism and nutrient-induced insulin secretion. *J. Biol. Chem.* 275, 35953–35959. doi:10.1074/jbc.M006612200

Wiggans, G. R., Vanraden, P. M., and Cooper, T. A. (2011). The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* 94, 3202–3211. doi:10.3168/jds.2010-3866

Wiggans, G. R., Cole, J. B., Hubbard, S. M., and Sonstegard, T. S. (2017). Genomic selection in dairy cattle: the USDA experience. *Annu. Rev. Anim. Biosci.* 5, 309–327. doi:10.1146/annurev-animal-021815-111422

Xu, D., Wang, Z., Xia, Y., Shao, F., Xia, W., Wei, Y., et al. (2020). The gluconeogenic enzyme PCK1 phosphorylates INSIG1/2 for lipogenesis. *Nature* 580, 530–535. doi:10.1038/s41586-020-2183-2

Yan, W., Zhou, H., Hu, J., Luo, Y., and Hickford, J. G. H. (2018). Variation in the FABP4 gene affects carcass and growth traits in sheep. *Meat Sci.* 145, 334–339. doi:10.1016/j.meatsci.2018.07.007

Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., et al. (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS One* 9, e93017. doi:10.1371/journal.pone.0093017

Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., et al. (2015). Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3 (Bethesda)* 5, 615–627. doi:10.1534/g3.114.016261

Zhenilo, S., Deyev, I., Litvinova, E., Zhigalova, N., Kaplun, D., Sokolov, A., et al. (2018). DeSUMOylation switches Kaiso from activator to repressor upon hyperosmotic stress. *Cell Death Differ.* 25, 1938–1951. doi:10.1038/s41418-018-0078-7

Zwierzchowski, L., Ostrowska, M., Zelazowska, B., and Bagnicka, E. (2021). Single nucleotide polymorphisms in the bovine SLC2A12 and SLC5A1 glucose transporter genes - the effect on gene expression and milk traits of Holstein Friesian cows. *Anim. Biotechnol.* 6, 1–11. doi:10.1080/10495398.2021.1954934

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/fgene.2022.994806

Check for updates

OPEN ACCESS

EDITED BY
Hongyu Liu,
Anhui Agricultural University, China

REVIEWED BY
Mao Yongjiang,
Yangzhou University, China
Da-Wei Wei,
Ningxia University, China

*CORRESPONDENCE
Zhidong Zhao,
zhaozd@gsau.edu.cn

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

RECEIVED 15 July 2022
ACCEPTED 06 September 2022
PUBLISHED 26 September 2022

CITATION
Li X, Bai Y, Li J, Chen Z, Ma Y, Shi B, Han X, Luo Y, Hu J, Wang J, Liu X, Li S and Zhao Z (2022), Transcriptional analysis of microRNAs related to unsaturated fatty acid synthesis by interfering bovine adipocyte ACSL1 gene.
Front. Genet. 13:994806.
doi: 10.3389/fgene.2022.994806

COPYRIGHT
© 2022 Li, Bai, Li, Chen, Ma, Shi, Han, Luo, Hu, Wang, Liu, Li and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Transcriptional analysis of microRNAs related to unsaturated fatty acid synthesis by interfering bovine adipocyte ACSL1 gene

Xupeng Li, Yanbin Bai, Jingsheng Li, Zongchang Chen, Yong Ma, Bingang Shi, Xiangmin Han, Yuzhu Luo, Jiang Hu, Jiqing Wang, Xiu Liu, Shaobin Li and Zhidong Zhao*

Gansu Key Laboratory of Herbivorous Animal Biotechnology, College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China

Long-chain fatty acyl-CoA synthase 1 (ACSL1) plays a vital role in the synthesis and metabolism of fatty acids. The proportion of highly unsaturated fatty acids in beef not only affects the flavor and improves the meat's nutritional value. In this study, si-ACSL1 and NC-ACSL1 were transfected in bovine preadipocytes, respectively, collected cells were isolated on the fourth day of induction, and then RNA-Seq technology was used to screen miRNAs related to unsaturated fatty acid synthesis. A total of 1,075 miRNAs were characterized as differentially expressed miRNAs (DE-miRNAs), of which the expressions of 16 miRNAs were upregulated, and that of 12 were downregulated. Gene ontology analysis indicated that the target genes of DE-miRNAs were mainly involved in biological regulation and metabolic processes. Additionally, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis identified that the target genes of DE-miRNAs were mainly enriched in metabolic pathways, fatty acid metabolism, PI3K-Akt signaling pathway, glycerophospholipid metabolism, fatty acid elongation, and glucagon signaling pathway. Combined with the previous mRNA sequencing results, several key miRNA-mRNA targeting relationship pairs, i.e., novel-m0035-5p—ACSL1, novel-m0035-5p—ELOVL4, miR-9-X—ACSL1, bta-miR-677—ACSL1, miR-129-X—ELOVL4, and bta-miR-485—FADS2 were screened via the miRNA-mRNA interaction network. Thus, the results of this study provide a theoretical basis for further research on miRNA regulation of unsaturated fatty acid synthesis in bovine adipocytes.

KEYWORDS

micrornas, ACSL1, bovine adipocytes, unsaturated fatty acids, RNA-seq

Introduction

Fat deposits and unsaturated fatty acid (UFA) content in beef muscle not only affect meat quality and flavor but also help improve the nutritional value of meat. Polyunsaturated fatty acids (PUFAs) play a significant role in chronic diseases, including cardiovascular disorders, cancers, and diabetes mellitus (Sijben and Calder, 2007; Marventano et al., 2015). Such as, long-chain n−3 polyunsaturated fatty acids, eicosapentaenoic acid (EPA, 20:5n−3) and docosahexaenoic acid (DHA; 22:6n−3) can reduce cardiovascular disease, the risks of cancer and type 2 diabetes have been widely recognized (Simopoulos, 1991; Barcelo-Coblijn and Murphy, 2009; Lopez-Huertas, 2010). Long-chain fatty acyl-CoA synthetases (ACSLs) are essential for fatty acid (FA) activation and catalysis. *ACSL1* converts long-chain FAs to fatty acyl-CoA (Paul et al., 2014). Previous studies have shown that *ACSL1* plays an important role in the activation of fatty acid synthesis of triglyceride (Li et al., 2009). In addition, *ACSL1* gene variants are also known to affect the content of unsaturated omega-3 FA, PUFAs, long-chain omega-3 FA, and docosapentaenoic acid in bovine skeletal muscle (Widmann et al., 2011). The results of follow-up studies have also shown that interference with the *ACSL1* gene caused reduced levels of MUFAs and PUFAs in bovine adipocytes. In contrast, the overexpression of the *ACSL1* gene was found to significantly upregulate the content of PUFAs (Tian et al., 2020; Zhao et al., 2020). Related studies have also found that miRNAs regulate lipid metabolism in milk (Lian et al., 2016), liver cancer cells (Cui et al., 2014), laying hens (Tian et al., 2019), and pig subcutaneous fat (Shan et al., 2022) by targeting *ACSL1*. Based on these results, *ACSL1* plays a key regulatory role in the synthesis of unsaturated fatty acids.

MicroRNAs (miRNAs) are endogenous non-coding small RNAs (approximately 22 nt long), which are widely found in eukaryotic cells (Shukla et al., 2011). They bind to the specific complementary site of the target mRNA at the post-transcriptional level, thereby degrading the mRNA or inhibiting its translation and regulating the protein expression (Bartel, 2009). Previous studies found that miRNAs play an important regulatory role in cell proliferation and differentiation, early animal development, apoptosis, oncogene expression inhibition, and fat metabolism (Tufekci et al., 2014; Deb et al., 2018). Many miRNAs are known to regulate the differentiation and deposition of adipocytes by regulating *PPAR*, C/EBPs, and other transcription factor families (Shukla et al., 2011). For example, bta-miR-130a/b is known to affect the differentiation of bovine adipocytes by targeting *PPARG* and *CYP2U1* (Ma et al., 2018). Also, miR-27a inhibits the differentiation of sheep preadipocytes by specifically binding to *RXRα* 3′UTR (Deng et al., 2020). Han et al. also found that miR-193a-5p inhibited the proliferation and differentiation of sheep preadipocytes by targeting the *ACAA2* gene (Han et al., 2021). A recent study found that miRNAs were involved in the cascade of genes and transcription factors to regulate fatty acid desaturation and fat deposition (Ren et al., 2020). These studies indicated that miRNAs exhibited a critical regulatory effect on fat metabolism.

This study aimed to further explore the molecular mechanism of miRNAs regulating the synthesis of unsaturated fatty acids in bovine adipocytes. High-throughput sequencing was used to systematically identify miRNAs and mRNAs (Bai et al., 2021) that were differentially expressed in bovine adipocytes after interference with the *ACSL1* gene and linked the mRNAs related to fatty acid metabolism with miRNAs to construct a miRNA-mRNA interaction network. The results provide new ideas and directions for the regulation of unsaturated fatty acid metabolism in bovine adipocytes.

Materials and methods

Sample collection

This animal study was reviewed and approved by the Faculty Animal Policy and Welfare Committee of Gansu Agricultural University (Ethic approval file No. GSAU-Eth-AST-2021-25). Calves (aged 1 day) from the livestock ranch of Gansu Agricultural University (Lanzhou, China) were selected. After sacrificing the calf, the perirenal adipose tissue was immediately collected for the isolation of bovine preadipocytes. The specific method of cell isolation is the same as our previous research (Tian et al., 2020).

Bovine pre-adipocyte culture and cell transfection

The bovine preadipocytes were cultured to the F3 generation, and then differentiation was induced *in vitro*. The test operation process and method followed the methods used in our previous studies (Tian et al., 2020). Finally, the cells on the fourth day of differentiation were collected in a 1.5 ml centrifuge tube without ribonuclease and were rapidly frozen in liquid nitrogen for subsequent RNA sequence analysis.

Construction and sequencing of small ribonucleic acid library

Total RNA was extracted from six adipocyte samples (3 with *ACSL1* gene interference, si group; the other three without *ACSL1* gene interference, NC group) using the Trizol kit (Invitrogen, Carlsbad, CA, USA). RNA degradation and contamination were analyzed via 1% agarose gel electrophoresis; RNA purity (OD260/280 ratio) was determined by Nanodrop analysis; RNA concentration was accurately quantified via Qubit; RNA

integrity was determined using Agilent 2,100. After quantification, the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA) was used to build the library. Using total RNA as the starting sample, polyacrylamide gel (PAGE) electrophoresis was run, select bands in the range of 18–30 nt were isolated, and recovers small RNA. Connect the 3′ adaptor and the 5' adaptor, respectively, and then perform reverse transcription and PCR amplification on the small RNAs connected to the adaptors on both sides. Finally, the PAGE gel was used to recover and purify the 140 bp band and dissolve it in EB solution to complete the library construction. The constructed library was tested for quality and yield using Agilent 2,100 and ABI StepOnePlus Real-Time PCR System (Life Technologies). After passing the quality inspection, the sequencing was performed by Novogene Bioinformatics Institute (Beijing, China) on Illumina HiSeq 2,500.

Identification of differential miRNAs

Raw reads were further filtered according to the following rules: 1) low quality reads containing more than one low quality (Q-value≤20) base or containing unknown nucleotides (N) were removed; 2) reads without 3′ adapters were removed; 3) reads containing 5′ adapters were removed; 4) reads containing 3′ and 5' adapters but no small RNA fragment between them were removed; 5) reads containing ployA in small RNA fragment were removed. After the quality control process, the clean tag sequence was aligned to the reference genome of the known cattle (Bos_taurus_Ensembl_94), and classification was performed by aligning it with the GeneBank (version 209.0), Rfam (11.0), and miRBase (22.0) databases. The sequence matching miRBase was considered as the known miRNA. According to software mirdeep2, new miRNA candidates were identified. The miRNA expression level was calculated and normalized to transcripts per million (TPM). For differentially expressed miRNA analysis, DESeq2 (V1.20.0) software was used. The screening criteria for differential miRNAs were the | fold change | > 2.0 and p-value < 0.05. The differential expression of all miRNAs, existing miRNAs, known miRNAs, and new miRNAs were analyzed simultaneously.

Functional enrichment analysis of differential miRNAs

Miranda (v3.3a) and TargetScan (Version: 7.0), two prediction software, were used to predict the target gene of miRNA, and the intersection of the target gene prediction results was used as the result of miRNA target gene prediction. The miRNA target genes were mapped to each term of the GO database (http://www.geneontology.org/), and the number of miRNA target genes was calculated for each term to obtain a list of miRNA target genes with a GO function miRNA target statistics of the number of genes. The hypergeometric test was used to find GO entries that were significantly enriched in miRNA target genes compared with the background. KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis used hypergeometric testing to find pathways that were significantly enriched in miRNA target genes compared with the entire background. Pathway significant enrichment was used to determine the most important biochemical metabolic pathways and signal transduction pathways involved in miRNA target genes. Finally, GO and KEGG pathways with Q-value < 0.05 were selected as the ones that were significantly enriched.

Real-time PCR verification

Ten miRNAs were selected for qRT-PCR to verify the differential expression results of sequencing. For miRNA expression, a miRNA first-strand cDNA synthesis kit (Accurate Biology, Hunan, China) was used to perform real-time fluorescent quantitative PCR. U6 snRNA was used as an internal reference. All qRT-PCR reactions were performed in the ABI 7500 real-time PCR system (Applied Biosystems, California, USA), with three reactions/samples. The relative expression of miRNA was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Construction of miRNA-mRNA interaction network

We selected nine signaling pathways for miRNA target gene enrichment, which are related to lipid metabolism. Then we intersect the genes enriched in the signaling path with the mRNAs expressed differently in our previous sequencing results (Bai et al., 2021), and use these mRNAs and miRNAs to build a miRNA-mRNA interaction network.

Results

Overview of small RNA sequencing

After high-throughput sequencing of six bovine adipocytes, the average clean reads of NC and si were 11, 368, 893 and 11,869,071, respectively, and the average clean tags obtained after quality control were 10,845,625 (95.40%) and 11,302,810 (95.23%) (Table 1). The raw reads obtained in the study were submitted to GenBank Temporary Submission ID: SUB11745359. The length of the smallest RNAs was in the range of 18–23 nt. The miRNAs of 22 nt were the longest, followed by miRNAs with a length of 20 nt, 21 nt, and 23 nt (Figure 1A). After comparing with GenBank, Rfam, and the reference genome, 93.59% of clean reads were identified as miRNA, and the remaining part included rRNA, scRNA, snRNA, snoRNA, tRNA, miRNA editing, and unann (Figure 1B). Approximately 96.15% of the tags were aligned to the reference

TABLE 1 Overview of small RNA sequencing.

Id	clean_reads	high_quality	3'adapter_null	insert_null	5'adapter_contaminants	polyA	clean_tags
NC-1	11,655,387 (100%)	11,568,102 (99.2511%)	26,054 (0.2252%)	49,674 (0.4294%)	14,982 (0.1295%)	103 (0.0009%)	10,986,690 (94.2628%)
NC-2	11,231,395 (100%)	11,097,944 (98.8118%)	59,021 (0.5318%)	4,456 (0.0402%)	3,871 (0.0349%)	112 (0.0010%)	10,857,341 (96.6696%)
NC-3	11,219,898 (100%)	11,079,983 (98.7530%)	33,458 (0.3020%)	8,190 (0.0739%)	7,796 (0.0704%)	110 (0.0010%)	10,692,844 (95.3025%)
si-1	11,463,101 (100%)	11,357,344 (99.0774%)	25,051 (0.2206%)	96,814 (0.8524%)	21,465 (0.1890%)	145 (0.0013%)	1,0,491,794 (91.5267%)
si-2	12,463,501 (100%)	1,2,301,997 (98.7042%)	48,443 (0.3938%)	6,041 (0.0491%)	4,318 (0.0351%)	125 (0.0010%)	12,072,144 (96.8600%)
si-3	11,680,612 (100%)	11,550,211 (98.8836%)	58,475 (0.5063%)	4,189 (0.0363%)	3,364 (0.0291%)	84 (0.0007%)	11,344,492 (97.1224%)

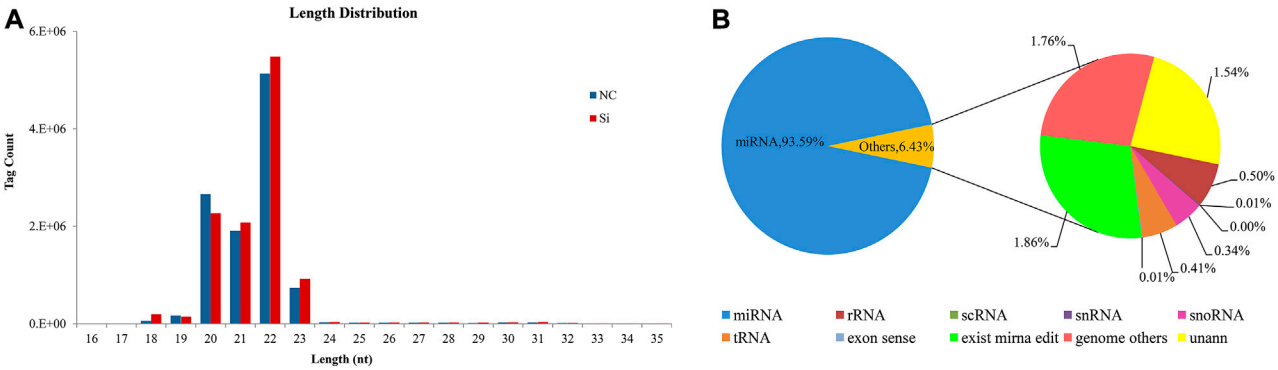


FIGURE 1
Correlation analysis of small RNA sequencing data. **(A)** Statistical analysis of small RNA fragment size. **(B)** Statistical analysis of the types of small RNA fragments after comparison with the database, including miRNA (existing miRNAs, known miRNAs and novel miRNAs), rRNA, scRNA, snRNA, snoRNA, tRNA, exon sense, miRNA editing, other genome, and unann.

genome (Bos_taurus_Ensembl_94). There were 491 bovine miRNAs and 462 known miRNAs. In addition, 122 new miRNAs were also predicted (Supplementary File S1).

Analysis of differentially expressed miRNAs

The DESeq2 (V1.20.0) software was used to compare the miRNAs between the si group and the NC group, and 28 DE-miRNAs were obtained, including 25 known and three new miRNAs. Among them, 16 upregulated miRNAs and 12 downregulated miRNAs were in the si group (Figure 2A). In general, the difference between the interference *ACSL1* gene (si group) and the control group (NC group) was highly correlated with the expression of these miRNAs. The clustering patterns of these 28 DE-miRNAs are shown in (Figure 2B).

GO and KEGG enrichment analysis of differentially expressed miRNA

TargetScan (Version: 7.0) and miRanda (v3.3a) was used to predict DE-miRNA target genes to determine their biological functions. The 28 DE-miRNAs predicted a total of 6,828 target genes (Supplementary File S2). GO enrichment analysis showed that these target genes participated in 1,111 significantly enriched functional classifications (Q-value < 0.05). Biological processes contained the most enriched genes with 817 GO terms, followed by cell components with 203 GO terms and molecular functions with 91 GO terms (Supplementary File S3). The enriched GO terms were mainly related to cellular processes, biological regulation, metabolic processes, cellular parts, binding, and catalytic activity (Figure 3A). KEGG results indicated that the target

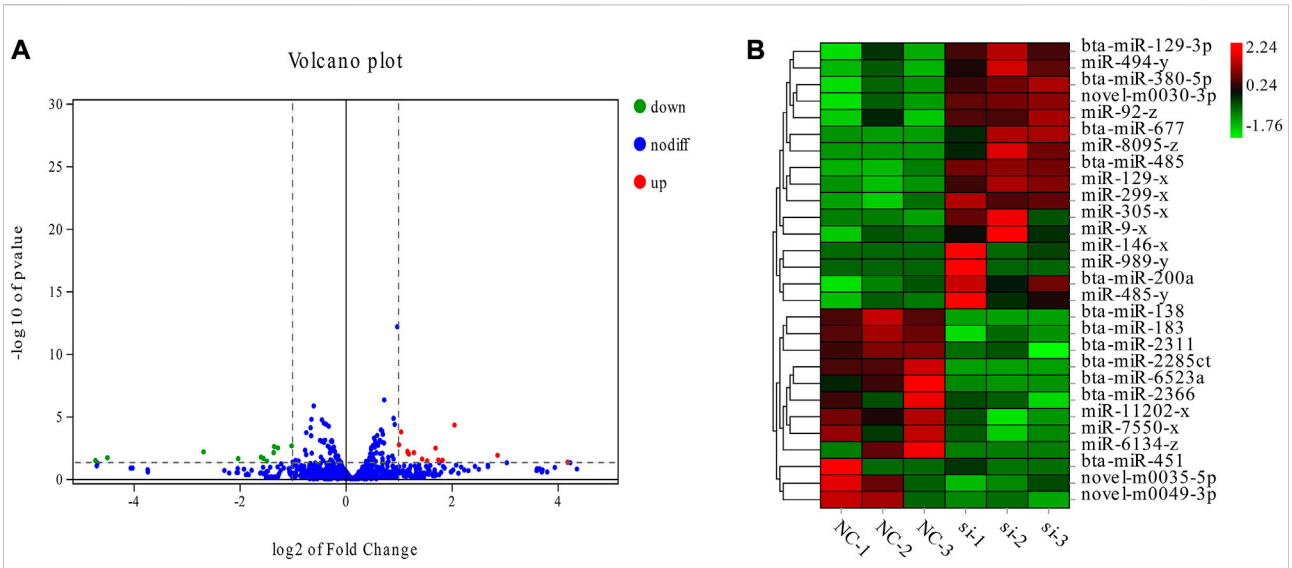


FIGURE 2
Statistical analysis of DE-miRNA. **(A)** The expression of miRNA volcano. The green dots on the left represent miRNAs that are significantly downregulated; the blue dots represent miRNAs that were not significantly different; and the red dots on the right represent miRNAs that were significantly upregulated. **(B)** Cluster map of differentially expressed miRNAs. Red indicates elevated expression and green indicates downregulated expression.

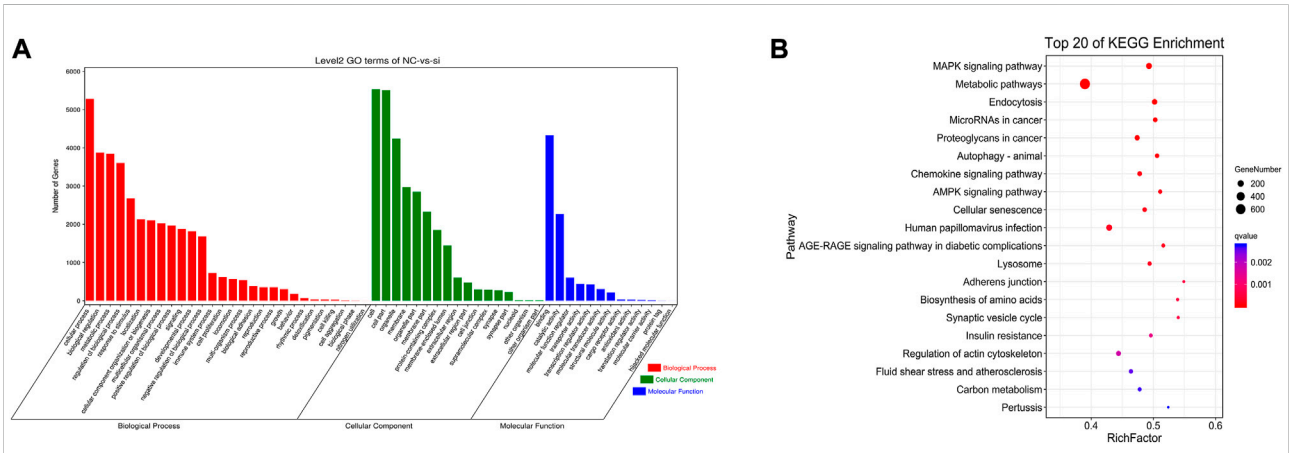


FIGURE 3
Functional enrichment analysis of DE-miRNAs. **(A)** GO enrichment analysis of target genes of DE-miRNAs. **(B)** Top 20 KEGG signaling pathways enriched by DE-miRNAs target genes.

genes of DE-miRNAs were significantly enriched in 81 signaling pathways (Q-value < 0.05) (Supplementary File S4). The enriched pathways included MAPK signaling pathway, metabolic pathway, AMPK signaling pathway, fatty acid metabolism, PI3K-Akt signaling pathway, glycerophospholipid metabolism, fatty acid elongation, steroid biosynthesis, and glucagon signaling pathway (Figure 3B).

Validation of differentially expressed miRNAs by qRT-PCR

The relative expression of 10 miRNAs was quantified by qRT-PCR to verify the differentially expressed miRNA (Figure 4). All selected DE-miRNAs showed consistent expression patterns between RNA-seq and qRT-PCR results, confirming the sequencing results.

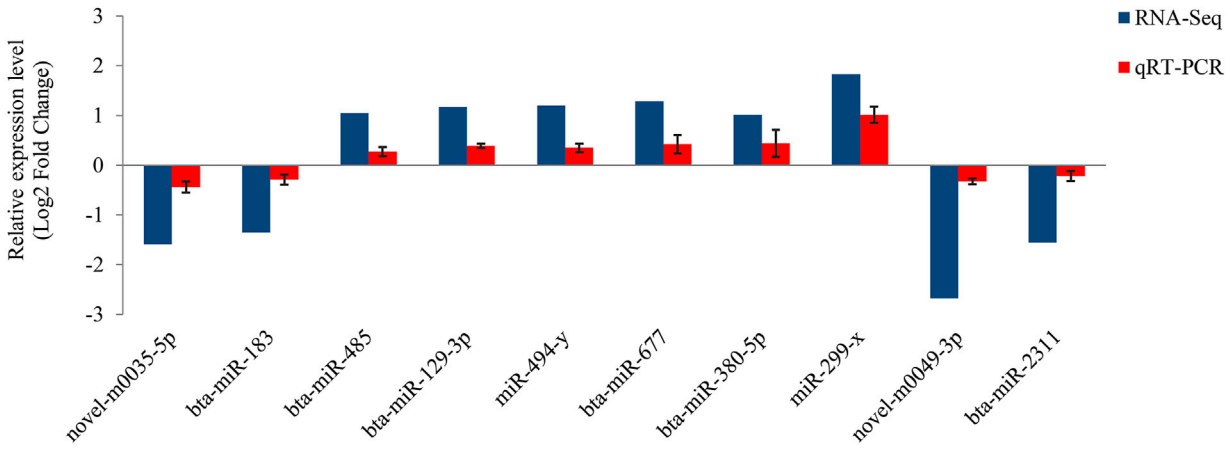


FIGURE 4
Validation of the miRNA-Seq results. The qRT-PCR validation was performed on 10 DE-miRNAs. These data show the mean ± SD for three replicates. Error bars indicate the standard deviation.

Construction of the miRNA-mRNA interaction network

The background genes related to fatty acid metabolism signaling pathways were screened to further understand and visualize the interaction between fat-related differentially expressed mRNAs (DEMs) and differentially expressed miRNAs (DE-miRNAs) (Figure 5A). Then these genes were used as target genes to construct a miRNA-mRNA interaction network (Figure 5B). In the network, 60 interactions were identified, and several of the most important DE-miRNAs were novel-m0035-5p, bta-miR-485, bta-miR-200a, bta-miR-677, miR-9-x, and miR-129-x.

Discussion

Fat deposition in muscle involves a series of processes, such as preadipocyte proliferation, preadipocyte differentiation, and lipid metabolism, and is regulated by related functional genes (Cai et al., 2018). MicroRNAs, as an important regulatory element after gene transcription, have been widely reported for their role in regulating lipid metabolism and synthesis (Du J et al., 2018; Jin et al., 2019; Khan et al., 2020; Zhang et al., 2020; Zhou et al., 2020). Therefore, it is necessary to understand the molecular mechanism of miRNAs in regulating the synthesis of unsaturated fatty acids (UFAs)-related genes to improve the nutritional value of beef.

In this study, 1,075 miRNAs were identified, of which 122 miRNAs were novel. Compared with the NC group, 28 miRNAs were identified as differentially expressed miRNAs. The DE-miRNAs included known lipid metabolism-related miRNAs such as miR-200a (Zhang et al., 2018), as well as newly discovered miRNAs novel-m0035-5p and novel-m0049-3p. GO and KEGG pathway enrichment analyses of miRNA target genes were carried out to reveal the potential function of DE-miRNAs. GO enrichment results showed that DE-miRNAs were mainly involved in cell processes, biological regulation, metabolic process, cellular parts, and binding. The results of KEGG analysis showed that DE-miRNAs target genes were mainly significantly enriched in 81 signal pathways ($Q < 0.05$). Nine pathways (Table 2) related to the synthesis of UFAs were selected for the follow-up research. The PI3K-Akt signaling pathway is a signaling pathway related to insulin (Pessin and Saltiel, 2000). Insulin regulates lipogenesis, fatty acid oxidation, VLDL-TG assembly, and secretion in goose liver cells through PI3K-Akt-mTOR to mediate lipid deposition (Han et al., 2015). Also, insulin and other hormones, such as glucagon, coordinate and regulate multiple biological responses. Insulin stimulates glucose transport in surrounding tissues, such as muscle and adipose tissue, inhibits glycogen synthesis and gluconeogenesis in the liver, and stimulates protein synthesis and lipogenesis (Wang and Sul, 1998). The MAPK signaling pathway can control biological processes via various cellular mechanisms involving the activation/inhibition of related factors (Yue and Lopez, 2020). Studies have found that miR-145 inhibits the lipid production of bovine preadipocytes by reducing the activity of PI3K/Akt and MAPK signaling pathways (Wang et al., 2020). Wu et al. found that miR-29a targeting *CTRP6* inhibited the proliferation of pig muscle and subcutaneous adipocytes via the MAPK signaling pathway but promoteed differentiation (Wu et al., 2017). AMPK is one of the important enzymes that regulate cell energy

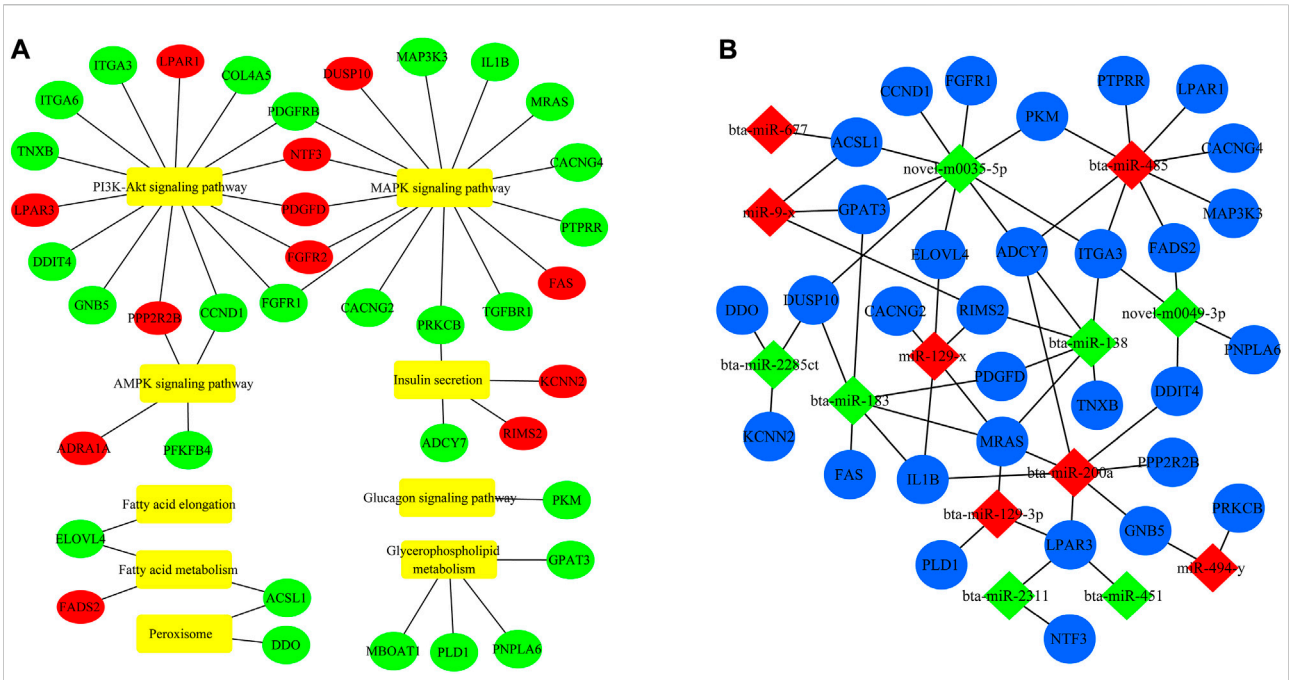


FIGURE 5
Interaction network of DE-miRNAs. **(A)** Network analysis of DE-miRNA target genes and their enrichment pathways. Red shows the upregulated mRNA. Green represents downregulated mRNA, and yellow represents the enriched signaling pathways. **(B)** Co-expression network of DE-miRNA and its target genes. Red is the miRNA whose expression is upregulated. Green represents downregulated miRNA, and blue represents mRNA.

metabolism and participates in cell metabolism. Activating AMPK in 3T3-L1 cells can inhibit the expression of key genes *C/EBPβ*, *PPARγ*, and *C/EBPα* in adipogenesis (Lee et al., 2012). We also found that more genes related to fat differentiation, such as *PPARγ*, *FASN*, and *SCD5*, are enriched in the AMPK signaling pathway. The enzymes required for β-oxidation of fatty acyl-CoA are present in peroxisomes and mitochondria. These enzymes promote the β-oxidation of long-chain and ultra-long-chain fatty acids in peroxisomes (Latruffe et al., 2001). They may also play a key role in regulating cellular senescence in related biochemical processes (Titorenko and Terlecky, 2011). An miRNA-mRNA interaction network was constructed to screen out miRNAs and mRNAs that may be related to UFA synthesis more intuitively and accurately. Here nine signaling pathways related to lipid metabolism were selected and then the background genes in the pathways were intersected with the differential mRNA sequenced from the previous transcriptome (Bai et al., 2021). These intersecting mRNAs and differential miRNAs were used to construct an interaction network. In the interaction network, the novel-m0035-5p and bta-miR-485 were found to simultaneously target multiple differential mRNAs. Most of the research on miR-485 is related to human cancer. A recent report discovered that miR-485 affected the content of triglycerides (TGs) and cholesterol (CHOL) in milk fat by targeting the *DTX4* gene (Liu et al., 2021).

In addition, several pairs of miRNA-mRNA targeting relationships of interest were identified, such as novel-m0035-5p—ACSL1, novel-m0035-5p—ELOVL4, bta-miR-677—ACSL1, bta-miR-485—FADS2, miR-129-x—ELOVL4, and miR-9-x—ACSL1. Studies have reported that variants of human fatty acid desaturase 1 (*FADS1*, encoding delta-5 desaturase) and fatty acid desaturase 2 (*FADS2*, encoding delta-6 desaturase) are associated with PUFAs (PUFA) in the blood and long-chain (LC) PUFA levels have a strong correlation (Glaser et al., 2010). *FADS2* gene polymorphism affects the content of unsaturated fatty acids (Gol et al., 2018; Proskura et al., 2019). In addition, studies have found that > C20 PUFAs synthesized by *FADS2* play an important role in regulating liver triacylglycerol and cholesterol accumulation during PUFA deficiency (Hayashi et al., 2021). As the key rate-limiting enzyme for the synthesis of >20 PUFAs, *FADS2* plays an important role in regulating the content of arachidonic acid (Stoffel et al., 2014). Hepatic steatosis was observed in *FADS2*-knockout mice that were fed a diet containing C18 PUFA but no C20 PUFA, and hepatic steatosis was reduced by administering arachidonic acid. Related studies have found that interference or overexpression of the *ACSL1* gene affects the expression of the downstream gene *COX2* and consequently the content of arachidonic acid (Cao et al., 2020).

TABLE 2 Nine significantly enriched pathways related to lipid metabolism.

Pathway ID	Pathway term	Qvalue	Target gene list
ko04151	PI3K-Akt signaling pathway	2.39E-02	ANGPT2,COL1A2,COL9A1,FGFR1,PRKAA1,CDKN1A,FGF2,PRLR,COL4A1,COL4A5,LPAR1,CD19,MAPK3
			COL6A1,GNB3,LPAR5,FGFR2,G6PC,ITGA6,COL6A5,LPAR2,PPP2R1A,PPP2R2C,CREB1,THBS1
ko04010	MAPK signaling pathway	3.07E-08	ANGPT2,FGF18,MAP3K1,DUSP2,FGFR1,IKBKB, TGFB2,TEK,MAPK14,PPP3CB,DUSP10,FGF2,NF1,AKT3
			NLK,SRF,MAPK8,EFNA4,CACNB3,MET,RRAS2,TNF,IL1A,MAP3K4,AKT1,FAS
ko04152	AMPK signaling pathway	1.89E-04	PPP2CB,EIF4EBP1,SIRT1,PFKFB2,PRKAA1,CD36,AKT3,MAP3K7,CREB5,ADRA1A,SREBF1,ADIPOR2
			PPARGC1A,PPP2R1A,FASN,SCD5,CREB1,CCND1,HMGCR
ko00062	Fatty acid elongation	4.12E-02	ELOVL4,HACD4,ELOVL5,PPT2,HACD2,ACOT7,HSD17B12,HADHB, HADHA,ECHS1,ELOVL6,ELOVL1,HADH
			PPT1,HACD3,ACOT4
ko01212	Fatty acid metabolism	5.31E-03	ACSL1,HSD17B8,CBR4,ELOVL4,HACD4,ELOVL5,HACD2,EHHADH, ACADS,ACAT2,HADHB, ACADM,ELOV6
			MCAT, FASN,SCD5,FADS1,FADS2,ACADL,HSD17B4,HACD3
ko00564	Glycerophospholipid metabolism	4.12E-02	GNPAT, AGPAT5,ETNK2,PLPP1,DGKE, GPAT4,DGKB,GPD1,AGPAT1,PCYT1A,PLA2G15,PLD1,DGKA, PISD
			PLA2G4A,PLD3,LPGAT1,GPAM,PLA2G12A,DGKD, LPIN3,GPAT3
ko04911	Insulin secretion	4.25E-02	CAMK2G,ATP1A2,FXYD2,ADCY7,ADCY6,ADCYAP1R1,CREB5,KCNMB4,GCK,VAMP2,SLC2A2,CREB3L1
			CACNA1F,PRKACB, PRKCB,KCNN4,CAMK2D,STX1A,SLC2A1
ko04922	Glucagon signaling pathway	4.12E-02	SIRT1,PRKAA1,PPP3CB,SIK2,CAMK2G,PRKAG1,GCK,PPP3CC,SLC2A2,PRKACB, PHKG1,CALM3,SLC2A1
			PPARGC1A,CPT1B,CREB1,ITPR1,PKM,CALM1
ko04146	Peroxisome	3.51E-02	GNPAT, ACSL1,PEX1,PEX11A,IDH2,NUDT7,PEX19,PEX12,CROT,PEX6,DDO,ABCD1,FAR1,PEX7,DECR2
			PEX10,EHHADH,MVK,ABCD3, SLC25A17,ABCD4

In the study of *ACSL1* gene promoter transcription regulation, several important transcription factors, were identified among which *Sp1*, the main negative transcription regulator, significantly reduced the activity of the *ACSL1* promoter (Zhao et al., 2016). Interestingly, *Sp1* could bind to the promoter region of *FADS2* to increase the promoter activity of *FADS2* (Li et al., 2019). Thus, it was hypothesized that *ACSL1* acted as a regulatory switch between the *FADS2* and *COX2* genes, which worked together to maintain the homeostasis of arachidonic acid content in the body. Very long chain fatty acid extension-4 (ELOVL4) is a fatty acid condensing enzyme that mediates the biosynthesis of very long chain PUFAs (VLC-PUFA; ≥ C28) in a limited number of tissues (Agbaga et al., 2014). Previous studies have shown that *ELOVL4* is required for the synthesis of C28 and C30 saturated fatty acids (VLC-FA) and C28-C38 very long-chain PUFAs (VLC-PUFA) (Agbaga et al., 2008), which are present in retinal photoreceptors, skin, sperm, and testis, which play an important role in normal and long-term function (McMahon et al., 2011; Harkewicz et al., 2012). Similar to fatty acid synthesis, the activity of enzymes involved in fatty acid extension and desaturation appears to be regulated primarily at the transcriptional level rather than through post-translational protein modification (Guillou et al., 2010). Therefore, it is necessary to understand the regulation of *ELOVL4* by miRNAs as post-transcriptional regulators.

MicroRNAs bind to specific complementary sites of the target mRNA, thereby degrading mRNA or inhibiting its translation and regulating protein expression (Bartel, 2009). This study showed that novel-m0035-5p—ACSL1, novel-m0035-5p—ELOVL4, and bta-miR-485—FADS2 had the same differential trend. However, this did not imply that the key role of these miRNAs had been overlooked. The expression of miRNAs varied in different tissues and at different stages. In addition, miRNA coordinated biological functions by targeting genes. Therefore, the mutual targeting effects of different tissues and different stages were also different (Wienholds and Plasterk, 2005; Bushati and Cohen, 2007). In addition, *ACSL1*, *FADS2*, and *ELOVL4* were significantly enriched in fatty acid metabolism and fatty acid extension pathways. The three aspects of differential genes, signal pathways, and gene function were comprehensively considered to determine the miRNAs related to the synthesis of unsaturated fatty acids. Future studies would verify the functional regulation of unsaturated fatty acid synthesis by these miRNAs.

Conclusion

RNA-Seq technology was used to identify miRNAs associated with UFAs synthesis by interfering with and none-interfering the *ACSL1* gene. Functional enrichment results indicated that DE-miRNAs were mainly involved in unsaturated fatty acid synthesis-related signaling pathways. Through the miRNA-mRNA interaction network, several key miRNA-mRNA targeting relationships were screened, including novel-m0035-5p—ACSL1, novel-m0035-5p—ELOVL4, miR-9-x—ACSL1, bta-miR-677—ACSL1, miR-129-x—ELOVL4 and bta-miR-485—FADS2. These miRNAs might regulate unsaturated fatty acid synthesis in bovine adipocytes by targeting these genes.

Data availability statement

The data presented in the study are deposited in the NCBI repository, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA856126 and the accession number: PRJNA856126.

Ethics statement

This animal study was reviewed and approved by the Faculty Animal Policy and Welfare Committee of Gansu Agricultural University (Ethic approval file No. GSAU-Eth-AST-2021-25).

Author contributions

Conceptualization, XL and ZZ; methodology, XL, YB, BS, and ZZ; validation, XL, YB, JL, YM, and ZC; formal analysis, XL, BS, YB, and SL; investigation, ZZ and XL; resources, YL and JW; writing—original draft preparation, XL; writing—review and editing, XL and ZZ; supervision, ZZ and XH; project administration, ZZ; funding acquisition, ZZ and JH All authors have read and agreed to the published version of the manuscript. Authorship have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (NO. 31860631); Young Doctor Fund Project of Gansu Provincial Department of Education (2021QB-027); Beef Cattle Genome Research (GSAU-ZL2015-031-01).

Acknowledgments

We thank all the members of Gansu Key Laboratory of Herbivorous Animal Biotechnology of the College of Animal Science and Technology, Gansu Agricultural University who contributed their efforts to these experiments. We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this 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.

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.2022.994806/full#supplementary-material

SUPPLEMENTARY FILE S1
Details of differentially expressed miRNAs.

SUPPLEMENTARY FILE S2
Details of differentially expressed miRNAs target genes.

SUPPLEMENTARY FILE S3
GO analysis of differentially expressed miRNAs target genes.

SUPPLEMENTARY FILE S4
KEGG analysis of differentially expressed miRNAs target genes.

References

Agbaga, M. P., Brush, R. S., Mandal, M. N., Henry, K., Elliott, M. H., and Anderson, R. E. (2008). Role of Stargardt-3 macular dystrophy protein (ELOVL4) in the biosynthesis of very long chain fatty acids. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12843–12848. doi:10.1073/pnas.0802607105

Agbaga, M. P., Logan, S., Brush, R. S., and Anderson, R. E. (2014). Biosynthesis of very long-chain polyunsaturated fatty acids in hepatocytes expressing ELOVL4. *Adv. Exp. Med. Biol.* 801, 631–636. doi:10.1007/978-1-4614-3209-8_79

Bai, Y., Li, X., Chen, Z., Li, J., Tian, H., Ma, Y., et al. (2021). Interference with ACSL1 gene in bovine adipocytes: Transcriptome profiling of mRNA and lncRNA related to unsaturated fatty acid synthesis. *Front. Vet. Sci.* 8, 788316. doi:10.3389/fvets.2021.788316

Barcelo-Coblijn, G., and Murphy, E. J. (2009). Alpha-linolenic acid and its conversion to longer chain n-3 fatty acids: Benefits for human health and a role in maintaining tissue n-3 fatty acid levels. *Prog. Lipid Res.* 48, 355–374. doi:10.1016/j.plipres.2009.07.002

Bartel, D. P. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell* 136, 215–233. doi:10.1016/j.cell.2009.01.002

Bushati, N., and Cohen, S. M. (2007). microRNA functions. *Annu. Rev. Cell Dev. Biol.* 23, 175–205. doi:10.1146/annurev.cellbio.23.090506.123406

Cai, H., Li, M., Sun, X., Plath, M., Li, C., Lan, X., et al. (2018). Global transcriptome analysis during adipogenic differentiation and involvement of transthyretin gene in adipogenesis in cattle. *Front. Genet.* 9, 463. doi:10.3389/fgene.2018.00463

Cao, Y., Wang, S., Liu, S., Wang, Y., Jin, H., Ma, H., et al. (2020). Effects of long-chain fatty acyl-CoA synthetase 1 on diglyceride synthesis and arachidonic acid metabolism in sheep adipocytes. *Int. J. Mol. Sci.* 21, E2044. doi:10.3390/ijms21062044

Cui, M., Wang, Y., Sun, B., Xiao, Z., Ye, L., and Zhang, X. (2014). MiR-205 modulates abnormal lipid metabolism of hepatoma cells via targeting acyl-CoA synthetase long-chain family member 1 (ACSL1) mRNA. *Biochem. Biophys. Res. Commun.* 444, 270–275. doi:10.1016/j.bbrc.2014.01.051

Deb, B., Uddin, A., and Chakraborty, S. (2018). miRNAs and ovarian cancer: An overview. *J. Cell. Physiol.* 233, 3846–3854. doi:10.1002/jcp.26095

Deng, K., Ren, C., Fan, Y., Liu, Z., Zhang, G., Zhang, Y., et al. (2020). miR-27a is an important adipogenesis regulator associated with differential lipid accumulation between intramuscular and subcutaneous adipose tissues of sheep. *Domest. Anim. Endocrinol.* 71, 106393. doi:10.1016/j.domaniend.2019.106393

Du, J., Xu, Y., Zhang, P., Zhao, X., Gan, M., Li, Q., et al. (2018). MicroRNA-125a-5p affects adipocytes proliferation, differentiation and fatty acid composition of porcine intramuscular fat. *Int. J. Mol. Sci.* 19, E501. doi:10.3390/ijms19020501

Glaser, C., Heinrich, J., and Koletzko, B. (2010). Role of FADS1 and FADS2 polymorphisms in polyunsaturated fatty acid metabolism. *Metabolism.* 59, 993–999. doi:10.1016/j.metabol.2009.10.022

Gol, S., Pena, R. N., Rothschild, M. F., Tor, M., and Estany, J. (2018). A polymorphism in the fatty acid desaturase-2 gene is associated with the arachidonic acid metabolism in pigs. *Sci. Rep.* 8, 14336. doi:10.1038/s41598-018-32710-w

Guillou, H., Zadravec, D., Martin, P. G., and Jacobsson, A. (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog. Lipid Res.* 49, 186–199. doi:10.1016/j.plipres.2009.12.002

Han, C., Wei, S., He, F., Liu, D., Wan, H., Liu, H., et al. (2015). The regulation of lipid deposition by insulin in goose liver cells is mediated by the PI3K-AKT-mTOR signaling pathway. *PLoS One* 10, e0098759. doi:10.1371/journal.pone.0098759

Han, F., Zhou, L., Zhao, L., Wang, L., Liu, L., Li, H., et al. (2021). Identification of miRNA in sheep intramuscular fat and the role of miR-193a-5p in proliferation and differentiation of 3T3-L1. *Front. Genet.* 12, 633295. doi:10.3389/fgene.2021.633295

Harkewicz, R., Du, H., Tong, Z., Alkuraya, H., Bedell, M., Sun, W., et al. (2012). Essential role of ELOVL4 protein in very long chain fatty acid synthesis and retinal function. *J. Biol. Chem.* 287, 11469–11480. doi:10.1074/jbc.M111.256073

Hayashi, Y., Lee-Okada, H. C., Nakamura, E., Tada, N., Yokomizo, T., Fujiwara, Y., et al. (2021). Ablation of fatty acid desaturase 2 (FADS2) exacerbates hepatic triacylglycerol and cholesterol accumulation in polyunsaturated fatty acid-depleted mice. *FEBS Lett.* 595, 1920–1932. doi:10.1002/1873-3468.14134

Jin, Y., Wang, J., Zhang, M., Zhang, S., Lei, C., Chen, H., et al. (2019). Role of bta-miR-204 in the regulation of adipocyte proliferation, differentiation, and apoptosis. *J. Cell. Physiol.* 234, 11037–11046. doi:10.1002/jcp.27928

Khan, R., Raza, S., Junjvlieke, Z., Wang, X., Wang, H., Cheng, G., et al. (2020). Bta-miR-149-5p inhibits proliferation and differentiation of bovine adipocytes through targeting CRTCs at both transcriptional and post-transcriptional levels. *J. Cell. Physiol.* 235, 5796–5810. doi:10.1002/jcp.29513

Latruffe, N., Cherkaoui, M. M., Nicolas-Frances, V., Jannin, B., Clemencet, M. C., Hansmannel, F., et al. (2001). Peroxisome-proliferator-activated receptors as physiological sensors of fatty acid metabolism: Molecular regulation in peroxisomes. *Biochem. Soc. Trans.* 29, 305–309. doi:10.1042/0300-5127:0290305

Lee, S. K., Lee, J. O., Kim, J. H., Kim, N., You, G. Y., Moon, J. W., et al. (2012). Coenzyme Q10 increases the fatty acid oxidation through AMPK-mediated PPARα induction in 3T3-L1 preadipocytes. *Cell. Signal.* 24, 2329–2336. doi:10.1016/j.cellsig.2012.07.022

Li, L. O., Ellis, J. M., Paich, H. A., Wang, S., Gong, N., Altshuller, G., et al. (2009). Liver-specific loss of long chain acyl-CoA synthetase-1 decreases triacylglycerol synthesis and beta-oxidation and alters phospholipid fatty acid composition. *J. Biol. Chem.* 284, 27816–27826. doi:10.1074/jbc.M109.022467

Li, Y., Zhao, J., Dong, Y., Yin, Z., Li, Y., Liu, Y., et al. (2019). Sp1 is involved in vertebrate LC-PUFA biosynthesis by upregulating the expression of liver desaturase and elongase genes. *Int. J. Mol. Sci.* 20, E5066. doi:10.3390/ijms20205066

Lian, S., Guo, J. R., Nan, X. M., Ma, L., Loor, J. J., and Bu, D. P. (2016). MicroRNA Bta-miR-181a regulates the biosynthesis of bovine milk fat by targeting ACSL1. *J. Dairy Sci.* 99, 3916–3924. doi:10.3168/jds.2015-10484

Liu, J., Jiang, P., Iqbal, A., Ali, S., Gao, Z., Pan, Z., et al. (2021). MiR-485 targets the DTX4 gene to regulate milk fat synthesis in bovine mammary epithelial cells. *Sci. Rep.* 11, 7623. doi:10.1038/s41598-021-87139-5

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Lopez-Huertas, E. (2010). Health effects of oleic acid and long chain omega-3 fatty acids (EPA and DHA) enriched milks. A review of intervention studies. *Pharmacol. Res.* 61, 200–207. doi:10.1016/j.phrs.2009.10.007

Ma, X., Wei, D., Cheng, G., Li, S., Wang, L., Wang, Y., et al. (2018). Bta-miR-130a/b regulates preadipocyte differentiation by targeting PPARG and CYP2U1 in beef cattle. *Mol. Cell. Probes* 42, 10–17. doi:10.1016/j.mcp.2018.10.002

Marventano, S., Kolacz, P., Castellano, S., Galvano, F., Buscemi, S., Mistretta, A., et al. (2015). A review of recent evidence in human studies of n-3 and n-6 PUFA intake on cardiovascular disease, cancer, and depressive disorders: Does the ratio really matter? *Int. J. Food Sci. Nutr.* 66, 611–622. doi:10.3109/09637486.2015.1077790

McMahon, A., Butovich, I. A., and Kedzierski, W. (2011). Epidermal expression of an Elovl4 transgene rescues neonatal lethality of homozygous Stargardt disease-3 mice. *J. Lipid Res.* 52, 1128–1138. doi:10.1194/jlr.M014415

Paul, D. S., Grevengoed, T. J., Pascual, F., Ellis, J. M., Willis, M. S., and Coleman, R. A. (2014). Deficiency of cardiac Acyl-CoA synthetase-1 induces diastolic dysfunction, but pathologic hypertrophy is reversed by rapamycin. *Biochim. Biophys. Acta* 1841, 880–887. doi:10.1016/j.bbalip.2014.03.001

Pessin, J. E., and Saltiel, A. R. (2000). Signaling pathways in insulin action: Molecular targets of insulin resistance. *J. Clin. Invest.* 106, 165–169. doi:10.1172/JCI10582

Proskura, W. S., Liput, M., Zaborski, D., Sobek, Z., Yu, Y. H., Cheng, Y. H., et al. (2019). The effect of polymorphism in the FADS2 gene on the fatty acid composition of bovine milk. *Arch. Anim. Breed.* 62, 547–555. doi:10.5194/aab-62-547-2019

Ren, H., Xiao, W., Qin, X., Cai, G., Chen, H., Hua, Z., et al. (2020). Myostatin regulates fatty acid desaturation and fat deposition through MEF2C/miR222/SCD5 cascade in pigs. *Commun. Biol.* 3, 612. doi:10.1038/s42003-020-01348-8

Shan, B., Yan, M., Yang, K., Lin, W., Yan, J., Wei, S., et al. (2022). MiR-218-5p affects subcutaneous adipogenesis by targeting ACSL1, a novel candidate for pig fat deposition. *Genes (Basel)* 13, 260. doi:10.3390/genes13020260

Shukla, G. C., Singh, J., and Barik, S. (2011). MicroRNAs: Processing, maturation, target recognition and regulatory functions. *Mol. Cell. Pharmacol.* 3, 83–92.

Sijben, J. W., and Calder, P. C. (2007). Differential immunomodulation with long-chain n-3 PUFA in health and chronic disease. *Proc. Nutr. Soc.* 66, 237–259. doi:10.1017/S0029665107005472

Simopoulos, A. P. (1991). Omega-3 fatty acids in health and disease and in growth and development. *Am. J. Clin. Nutr.* 54, 438–463. doi:10.1093/ajcn/54.3.438

Stoffel, W., Hammels, I., Jenke, B., Binczek, E., Schmidt-Soltau, I., Brodesser, S., et al. (2014). Obesity resistance and deregulation of lipogenesis in Δ6-fatty acid desaturase (FADS2) deficiency. *EMBO Rep.* 15, 110–120. doi:10.1002/embr.201338041

Tian, H. S., Su, X. T., Han, X. M., Zan, L. S., H, J., Luo, Y. Z., et al. (2020). Effects of silencing ACSL1 gene by siRNA on the synthesis of unsaturated fatty acids in adipocytes of qinchuan beef cattle. *J. Agric. Biotechnol.* 10, 1722–1732. doi:10.3969/j.issn.1674-7968.2020.10.002

Tian, W. H., Wang, Z., Yue, Y. X., Li, H., Li, Z. J., Han, R. L., et al. (2019). miR-34a-5p increases hepatic triglycerides and total cholesterol levels by regulating ACSL1 protein expression in laying hens. *Int. J. Mol. Sci.* 20, E4420. doi:10.3390/ijms20184420

Titorenko, V. I., and Terlecky, S. R. (2011). Peroxisome metabolism and cellular aging. *Traffic* 12, 252–259. doi:10.1111/j.1600-0854.2010.01144.x

Tufekci, K. U., Meuwissen, R. L., and Genc, S. (2014). The role of microRNAs in biological processes. *Methods Mol. Biol.* 1107, 15–31. doi:10.1007/978-1-62703-748-8_2

Wang, D., and Sul, H. S. (1998). Insulin stimulation of the fatty acid synthase promoter is mediated by the phosphatidylinositol 3-kinase pathway. Involvement of protein kinase B/Akt. *J. Biol. Chem.* 273, 25420–25426. doi:10.1074/jbc.273.39.25420

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

Wienholds, E., and Plasterk, R. H. (2005). MicroRNA function in animal development. *FEBS Lett.* 579, 5911–5922. doi:10.1016/j.febslet.2005.07.070

Widmann, P., Nuernberg, K., Kuehn, C., and Weikard, R. (2011). Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. *BMC Genet.* 12, 96. doi:10.1186/1471-2156-12-96

Wu, W., Zhang, J., Zhao, C., Sun, Y., Pang, W., and Yang, G. (2017). CTRP6 regulates porcine adipocyte proliferation and differentiation by the AdipoR1/MAPK signaling pathway. *J. Agric. Food Chem.* 65, 5512–5522. doi:10.1021/acs.jafc.7b00594

Yue, J., and Lopez, J. M. (2020). Understanding MAPK signaling pathways in apoptosis. *Int. J. Mol. Sci.* 21, E2346. doi:10.3390/ijms21072346

Zhang, Q., Ma, X. F., Dong, M. Z., Tan, J., Zhang, J., Zhuang, L. K., et al. (2020). MiR-30b-5p regulates the lipid metabolism by targeting PPARGC1A in Huh-7 cell line. *Lipids Health Dis.* 19, 76. doi:10.1186/s12944-020-01261-3

Zhang, Y., Wu, X., Liang, C., Bao, P., Ding, X., Chu, M., et al. (2018). MicroRNA-200a regulates adipocyte differentiation in the domestic yak Bos grunniens. *Gene* 650, 41–48. doi:10.1016/j.gene.2018.01.054

Zhao, Z., Abbas, R. S., Tian, H., Shi, B., Luo, Y., Wang, J., et al. (2020). Effects of overexpression of ACSL1 gene on the synthesis of unsaturated fatty acids in adipocytes of bovine. *Arch. Biochem. Biophys.* 695, 108648. doi:10.1016/j.abb.2020.108648

Zhao, Z. D., Zan, L. S., Li, A. N., Cheng, G., Li, S. J., Zhang, Y. R., et al. (2016). Characterization of the promoter region of the bovine long-chain acyl-CoA synthetase 1 gene: Roles of E2F1, Sp1, KLF15, and E2F4. *Sci. Rep.* 6, 19661. doi:10.1038/srep19661

Zhou, X., Shi, X., Wang, J., Zhang, X., Xu, Y., Liu, Y., et al. (2020). miR-324-5p promotes adipocyte differentiation and lipid droplet accumulation by targeting Krueppel-like factor 3 (KLF3). *J. Cell. Physiol.* 235, 7484–7495. doi:10.1002/jcp.29652

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fgene.2022.1022261

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Zahoor Khan,
University of Agriculture, Dera Ismail Khan, Pakistan

REVIEWED BY
Serafino M. A. Augustino,
University of Juba, South Sudan
Xianyong Lan,
Northwest A&F University, China

*CORRESPONDENCE
Chonglong Wang,
ahwchl@163.com

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

RECEIVED 18 August 2022
ACCEPTED 28 September 2022
PUBLISHED 17 October 2022

CITATION
Zhang W, Li X, Jiang Y, Zhou M, Liu L, Su S, Xu C, Li X and Wang C (2022), Genetic architecture and selection of Anhui autochthonous pig population revealed by whole genome resequencing.
Front. Genet. 13:1022261.
doi: 10.3389/fgene.2022.1022261

COPYRIGHT
© 2022 Zhang, Li, Jiang, Zhou, Liu, Su, Xu, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic architecture and selection of Anhui autochthonous pig population revealed by whole genome resequencing

Wei Zhang, Xiaojin Li, Yao Jiang, Mei Zhou, Linqing Liu, Shiguang Su, Chengliang Xu, Xueting Li and Chonglong Wang*

Key Laboratory of Pig Molecular Quantitative Genetics of Anhui Academy of Agricultural Sciences, Anhui Provincial Key Laboratory of Livestock and Poultry Product Safety Engineering, Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei, China

The genetic resources among pigs in Anhui Province are diverse, but their value and potential have yet to be discovered. To illustrate the genetic diversity and population structure of the Anhui pigs population, we resequenced the genome of 150 pigs from six representative Anhui pigs populations and analyzed this data together with the sequencing data from 40 Asian wild boars and commercial pigs. Our results showed that Anhui pigs were divided into two distinct types based on ancestral descent: Wannan Spotted pig (WSP) and Wannan Black pig (WBP) origins from the same ancestor and the other four populations origins from another ancestor. We also identified several potential selective sweep regions associated with domestication characteristics among Anhui pigs, including reproduction-associated genes (CABS1, INSL6, MAP3K12, IGF1R, INSR, LIMK2, PATZ1, MAPK1), lipid- and meat-related genes (SNX19, MSTN, MC5R, PRKG1, CREBBP, ADCY9), and ear size genes (MSRB3 and SOX5). Therefore, these findings expand the catalogue and how these genetic differences among pigs and this newly generated data will be a valuable resource for future genetic studies and for improving genome-assisted breeding of pigs and other domesticated animals.

KEYWORDS

whole genome sequencing, SNP, signatures of selection, Anhui pig population, population structure

1 Introduction

Domestic pig (*Sus scrofa*) is an important livestock species, and has served as a source of stable food, organic fertilizer, industrial raw material, and medicine for humans since its domestication in the early Neolithic period. Since then, the domestic pig has promoted population growth and socioeconomic transformation from a hunter-gatherer society to sedentary agricultural settlements. Through domestication, natural

and artificial selection has generated 48 pig breeds in China (China National Commission of Animal Genetic Resources. 2011). Among evolutionary biologists, there is substantial interest in elucidating how natural and artificial selection have shaped modern pig genomes and the how these genetic differences contribute to the complex phenotypes seen among pigs.

Whole-genome resequencing has the potential to resolve pig population structure and identify phenotype-associated functional genes that arose during domestication and breeding of different pig breeds. In recent years, whole-genome sequencing studies on germplasm characteristics among different pig breeds have revealed the evolutionary history of the pig population and elucidated the genetic basis of complex phenotypes, while also providing new focuses in the future breeding of Chinese indigenous pigs (Rubin et al., 2012; Li et al., 2013; Ai et al., 2015; Frantz et al., 2015). Several causative genes have been reported as responsible for phenotypic variation among Chinese indigenous pigs. For instance, HIFA has been identified as the gene responsible for hypoxic adaptation in Tibetan pigs (Li et al., 2013). Moreover, TGFB3 and DAB2IP have been shown to be important in regulating the number of ribs (Zhu et al., 2015). Additionally, MITF, EDNRB, and MC1R have been identified as the genes responsible for color variations among Chinese breeds (Fang et al., 2009; Wang et al., 2015; Zhao et al., 2018). However, most of the aforementioned studies focused only on a few phenotypes among a limited number of breeds or populations. Therefore, considering that China has such a large and a diverse number of pig breeds, more genes with germplasm characteristics still need to be identified.

The Anhui Province, located in the Yangtze River Delta, is one of the important birthplaces of Chinese civilization. Additionally, the province is home to a wide variety of indigenous pig populations, including the Dingyuan pig (DYP), Huoshou Black pig (HBP), Wannan Spotted pig (WSP), Wannan Black pig (WBP), Anqing Six-end-white pig (ASP), and Wei pig (WP). These pig breeds are representative autochthonous Chinese breeds with long histories of breeding in the country. Of mention is that Anhui indigenous pig populations are facing crises of population decline and loss of genetic characteristics. The reasons for this are complex, including long cycles of feeding, the introduction of modern commercial pigs, etc. however, insufficient protection is the most direct cause. To date, there have been no systematic studies (i.e., based on a large panel of Anhui indigenous pig populations) seeking to illustrate the genomic diversity and signatures of selection among Anhui indigenous pig populations. Therefore, a comprehensive study of the genetic diversity, phylogenetic relationships, population structure, demographic history, and selection signatures among the Anhui pig populations are critical. Furthermore, the Anhui pig populations represent unique genetic resources for understanding the role of functional diversity, which can be applied in other contexts as well. Moreover, characterizing genetic variations and identifying the associated phenotypes are vital for future breeding of indigenous pig populations.

For this purpose, we performed whole-genome sequencing of six representative pig populations in Anhui (WSP, n = 25; WBP, n = 25; ASP, n = 25; WP, n = 25; DYP, n = 25; HBP, n = 25). Using whole-genome sequencing data in conjunction with downloaded sequence data, we explored the genetic diversity, phylogenetic relationships, population structure, and demographic history of the six pig populations. In addition, we calculated the F_{ST} (fixation index) and $\log_2$ ($\theta\pi$ ratio) values to elucidate the signatures of selection and then characterized these candidates' genes are present in the genome of Anhui pigs. Our findings provide valuable insights into the evolutionary history of Anhui pig populations and reveal several promising candidate genes for future indigenous pig breeding.

2 Materials and methods

2.1 Sample collection and genome sequencing

To survey the overall genetic diversity among autochthonous pig breeds in Anhui province, blood samples from 150 individuals within six indigenous Anhui pig populations were collected from pig conservation farms (Figure 1; Supplementary Table S1). These breeds included: ASP (N = 25, ♀15, ♂10), WSP (N = 25, ♀10, ♂15), WBP (N = 25, ♀11, ♂14), WP (N = 25, ♀13, ♂12), DYP (N = 25, ♀4, ♂21), HBP (N = 25, ♀10, ♂15). This study was conducted in accordance with and was approved by the Animal Care Committee of the Anhui Academy of Agricultural Sciences (Hefei, China; no. AAAS2020-04). For the blood samples, genomic DNA was extracted using the standard phenol–chloroform method (Sambrook and Russel 2001) and stored at 4°C to avoid freeze–thawing. The quality and concentration of the DNA were assessed using a 0.5% agarose gel (run for >8 h at 25 V) and a Nanodrop spectrophotometer (Thermo Fisher Scientific). DNA was then fragmented and treated in accordance with the Illumina DNA sample preparation protocol. Sequencing libraries were constructed and sequenced on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, United States) using paired-end 150 bp reads through the Novogene service (Beijing, China).

For comparison, 40 pig resequenced data sets from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/sra/) were downloaded (Ai et al., 2015; Rubin et al., 2012; Zhao et al., 2018; Groenen et al., 2012) (Supplementary Table S2), including Asian wild boar, Duroc, Pietrain, Yorkshire, and Landrace.

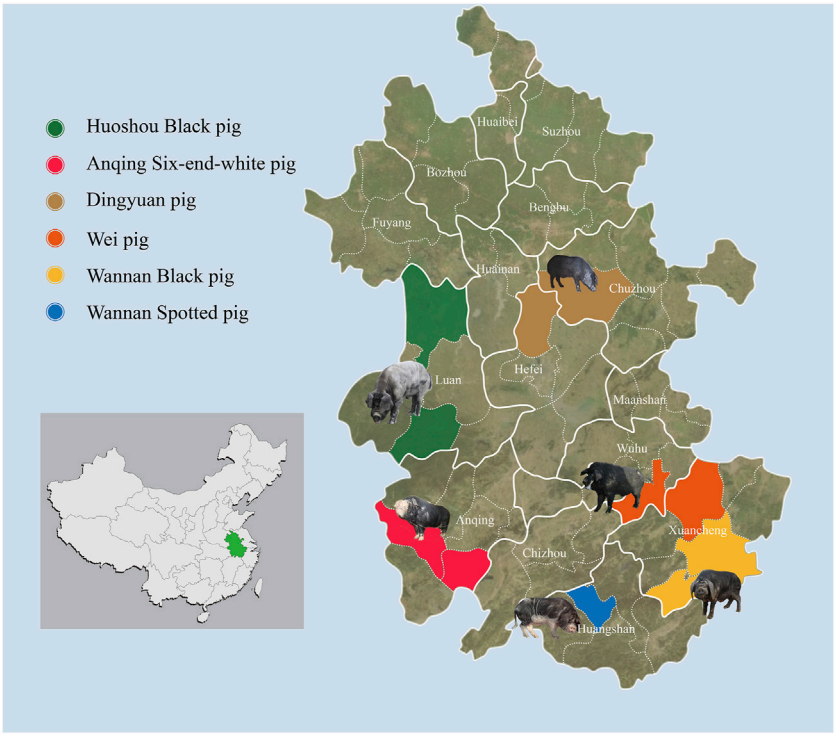


FIGURE 1
Geographic distribution and appearances of typical pigs. 150 Anhui indigenous pigs representing six populations were included.

2.2 Read mapping and variant calling

Before alignment, adapter and low-quality reads [e.g., reads with 10% unidentified nucleotides (*N*); reads with >10 nt aligned to the adapter; reads with >30% bases with Phred quality ≤25] were filtered using the NGSQCToolkit_v2.3.351. The remaining, high-quality reads were aligned with the pig reference genome (ftp://ftp.ensembl.org/pub/release-99/variation/gvf/sus_scrofa/) using BWA-0.7.12 software (https://sourceforge.net/projects/bio-bwa/) default parameters (Li and Durbin. 2009). For the mapping process, a BAM file index was built using SAMtools (https://github.com/samtools/samtools/releases/) (Li et al., 2009). The BAM file was sorted using the Picard-tools-1.105 software (https://github.com/broadinstitute/picard/releases).
Duplicate read data were excluded. Next, we performed variant calling for all samples on a population-scale using SAMtools mpileup with the parameters "-m 2 -F 0.002 -d 1000" and then bcftools to view variants (Li and Durbin 2011). We further resolved the SNPs according to the following criteria: QD (Variant Confidence/Quality by Depth) < 5.0, MQ (RMS Mapping Quality) < 40.0, FS (Phred-scaled *p* value using Fisher's exact test to detect strand bias) > 60.0, SOR (Strand Odd Ratio) > 3.0, MQRankSum (Z-score from Wilcoxon rank sum test of Alt vs. Ref read mapping qualities) < −10.0, ReadPosRankSum (Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias) < −8.0 and QUAL < 30. SNPs that were bi-allelic missed >40% of calls and had a MAF < 0.05 were removed and the remaining SNPs comprised the basic set. Variants were annotated using ANNOVAR (Wang et al., 2011). The downloaded resequencing data were analyzed using the process described above.

2.3 Genomic diversity and population genetic structure

Nucleotide diversity (θ_π) was calculated based on the maximum likelihood estimates of the fold-site frequency spectrum using a sliding window approach (50-kb windows with 50-kb steps; Kim et al., 2011). To illustrate the population structure, principal component analysis was performed using the GCTA software (v.1.25; Yang et al., 2011). Additionally, phylogenetic trees were inferred through the neighbor-joining (NJ) method and implemented in TreeBest (http://treesoft.sourceforge.net/treebest.shtml) with 1000 bootstraps. The population structure was deduced using the ADMIXTURE software (https://dalexander.github.io/admixture/index.html) with a kinship (K) set from 2 to 6 (Alexander et al., 2009).

2.4 Pairwise sequential Markovian coalescence analysis

The demographic history of the six Anhui indigenous pig populations was examined using the hidden Markov model approach and implemented according to the pairwise sequential Markovian coalescence (PSMC) model (Li and Durbin 2011). To estimate the distribution time, we used the parameters of $g = 5$ and a rate of 2.5×10^{-8} mutations per generation.

2.5 Selective sweeps analysis

To identify genetic variants associated with adaptation of the Anhui pig populations, comparisons were performed between three populations: 1) the Anhui pig populations versus the Asian wild boar, and 2) the Anhui pig populations versus the commercial pig population. The θ_π and F_{ST} were calculated using a 40-kb sliding window approach with a 20-kb step-size, in PopGenome (Pfeifer et al., 2014). The overlapping windows within the top 1% or 5% of the F_{ST} and $\theta\pi$ ratio ($\log_2$-transformed) distributions were considered putative selective regions (Li et al., 2013). Gene contents in "significant" genomic regions were retrieved from the Ensembl Genes 102 Database and analyzed in the BioMart software (http://asia.ensembl.org/biomart/martview/). To further explore the potential biological significance of the genes within these sweep regions, Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted to perform functional annotation for these genes using KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/). The terms and pathways exhibiting p-values < 0.05 were considered significant.

3 Results

3.1 Whole-genome sequencing and genetic variations

In total, 150 individuals from six pig populations indigenous to Anhui were selected for whole-genome resequencing. In total, 31 billion reads (or ~4,678 Gb of sequences) were generated. Average data sets consisted of 31.2 Gb (10.85 X) per individual and 800.36, 750.16, 724.8, 797.18, 834.25, and 771.53 Gb for the HBP, ASP, WSP, DYP, WBP, and WP collective populations, respectively. We also downloaded data for 40 non-indigenous individuals (Asian wild boars, Duroc, Landrace, Yorkshire, and Pitelan), resulting in 5,732 Gb of data in total. Using BWA, reads were aligned to the pig reference genome with an average alignment rate of 98.49% ± 0.14% and an estimated error rate of 0.03% ± 0.005% (Supplementary Tables S3, S4). After stringent variant calling and filtering, approximately 23 million single nucleotide polymorphisms (SNPs) were identified. For all detected SNPs, the average transition/transversion ratio was 2.42, which is consistent with a previous study (Kerstens et al., 2009). Further annotation of these SNPs revealed that they were most abundant in intergenic regions (53.58%) and intronic regions (42.97%), followed by downstream (0.62%) and upstream (0.56%) regions, whereas only 0.7% were located in coding sequences. Most variations were located in non-coding sequences, indicating that during domestication and breeding, the non-coding sequences could potentially change protein function by regulating gene expression.

3.2 Genomic diversity and population genetic structure

Nucleotide diversity reflects the degree of polymorphism within a population (Nei and Li. 1979). On a genome-wide window scale of 50-kb, with steps of 50-kb, the foreign pig breeds displayed reduced nucleotide diversity compared with the indigenous Anhui populations (Figure 2A). This is likely the result of extensive artificial selection over generations. We performed PCA analysis for the experimental populations in GCTA (Figure 2B). This analysis was successful in separating breed clusters based on genotypic data, where the foreign breeds, ASP and WP, HBP and DYP, and WBP and WSP were clustered together. These findings are consistent with the established information on the Anhui populations. To elucidate the phylogenetic relationship among the Anhui pig breeds, unrooted phylogenetic trees were constructed, revealing genetically distinct clusters (Figure 2C). This was consistent with the results of the PCA analysis, revealing clustering into distinct genetic groups. We used the ADMIXTURE software to determine the degree of admixture when K is increased from 2 to 6, where K is the assumed number of ancestral populations. $K = 2$ was suggested as the most plausible number among our samples (Figure 2D), reflecting the divergence of the Anhui populations and commercial breeds within the pig population. These results show that the current pig populations (WP, DYP, ASP, and HBP) may have some of the same genetic information as foreign pigs, possibly owing to the introduction of modern commercial pigs. We also conducted an admixture analysis of the Anhui populations (Figure 2E), and the results showed that $K = 2$ was the probable number of genetically distinct groups within the Anhui populations. It was revealed that WSP and WBP originated from the same ancestor and the other four populations from another ancestor. At $K = 3$, WBP was separated from WSP, and the others still descended from

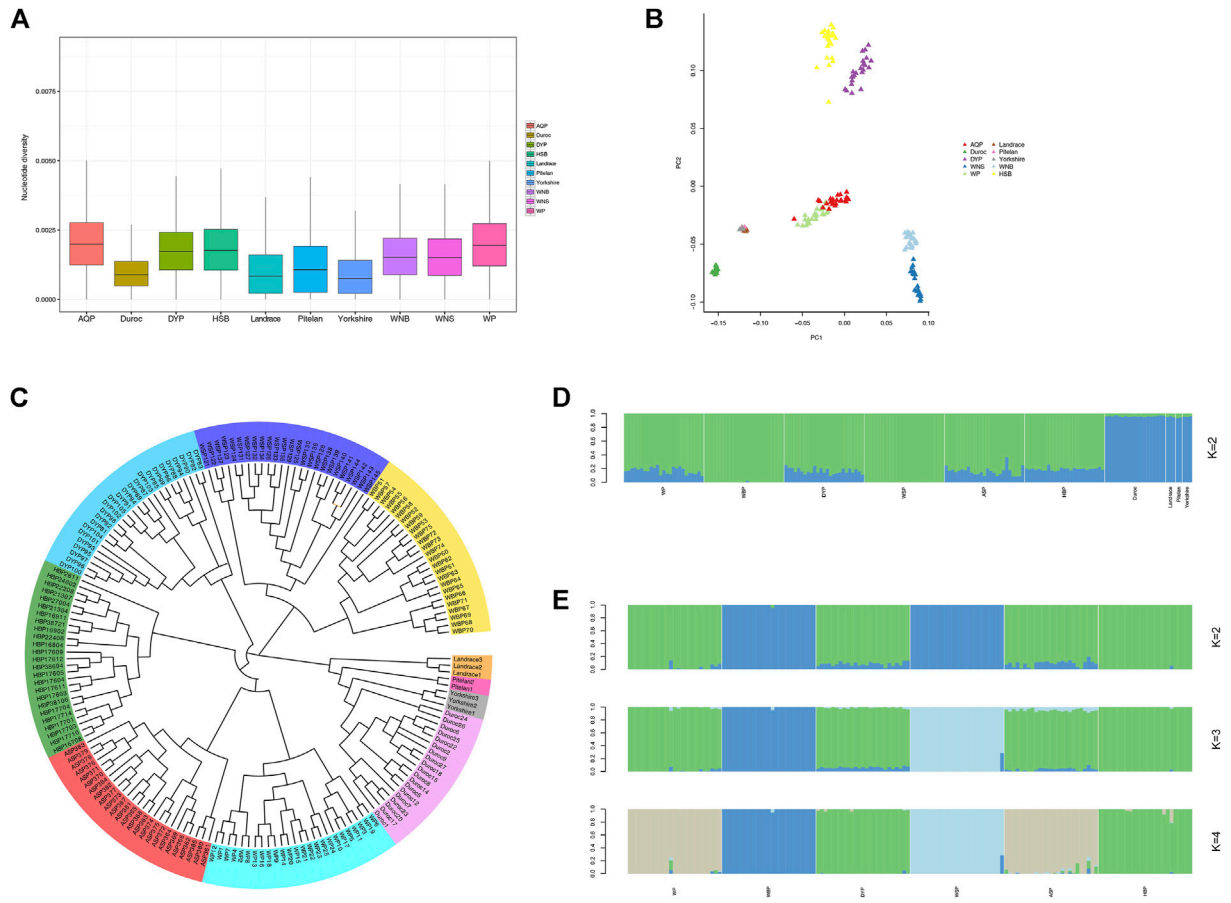


FIGURE 2
Population structure and relationship of Anhui in comparison to commercial pig. **(A)** Genome-wide distribution of nucleotide diversity among Anhui pig breeds and four commercial pig breeds with a 50-kb window. **(B)** Principal component analysis of Anhui pig breeds and four commercial pig breeds, PC1 against PC2. **(C)** Neighbor-joining tree of the relationships between the ten pig populations. The scale bar represents the identity-by-state (IBS) score between pairs of animals. **(D)** Proportion of ancestry for each individual assuming different number of ancestral population (*K* = 2). Colors in each vertical line represent the likelihood proportion of an animal genome assigned to a source population. **(E)** Proportion of ancestry for each individual in Anhui pig populations.

the same ancestor. At $K = 4$, WBP was still distinct from WSP, WP was grouped with ASP, and DYP was grouped with HBP, in agreement with prior knowledge of Anhui pig populations.

3.3 Demographic history

To explore the demographic histories of the six Anhui pig populations, PSMC was used to infer the historical population size. The variations in effective population size (Ne) over time are shown in Figure 3. Individuals from the same breed displayed similar historical fluctuations in effective population size. All breeds shared a collective population decline beginning at approximately 10,000 BP, likely a consequence of Neolithic domestication events (Zeder, 2008; Boitard et al., 2016).

3.3.1 The selective regions of Anhui pig populations compared to those of the Asian wild boar population

To identify the signatures of selection, we compared the genomes of Anhui pig populations to those of Asian wild boars. Specifically, we used both θπ- and F_{ST}-based cross approaches to investigate the signatures of selection. For this study, the regions within the top 1% were defined as the selected regions. The genome distribution of the two statistics were shown in Figures 4A,B. A total of 402 selective regions harboring 187 genes were identified in the top 1% of the F_{ST} and $\log_2$ (θπ, ratios) distributions (16.1 Mb of the genome, Figure 4C and Supplementary Tables S5, S6). All 187 genes were used for GO and KEGG analyses (Supplementary Tables S7, S8 and Supplementary Figures S1, S2). Functional annotation revealed that the selected genes may have effects on reproduction (CABS1,

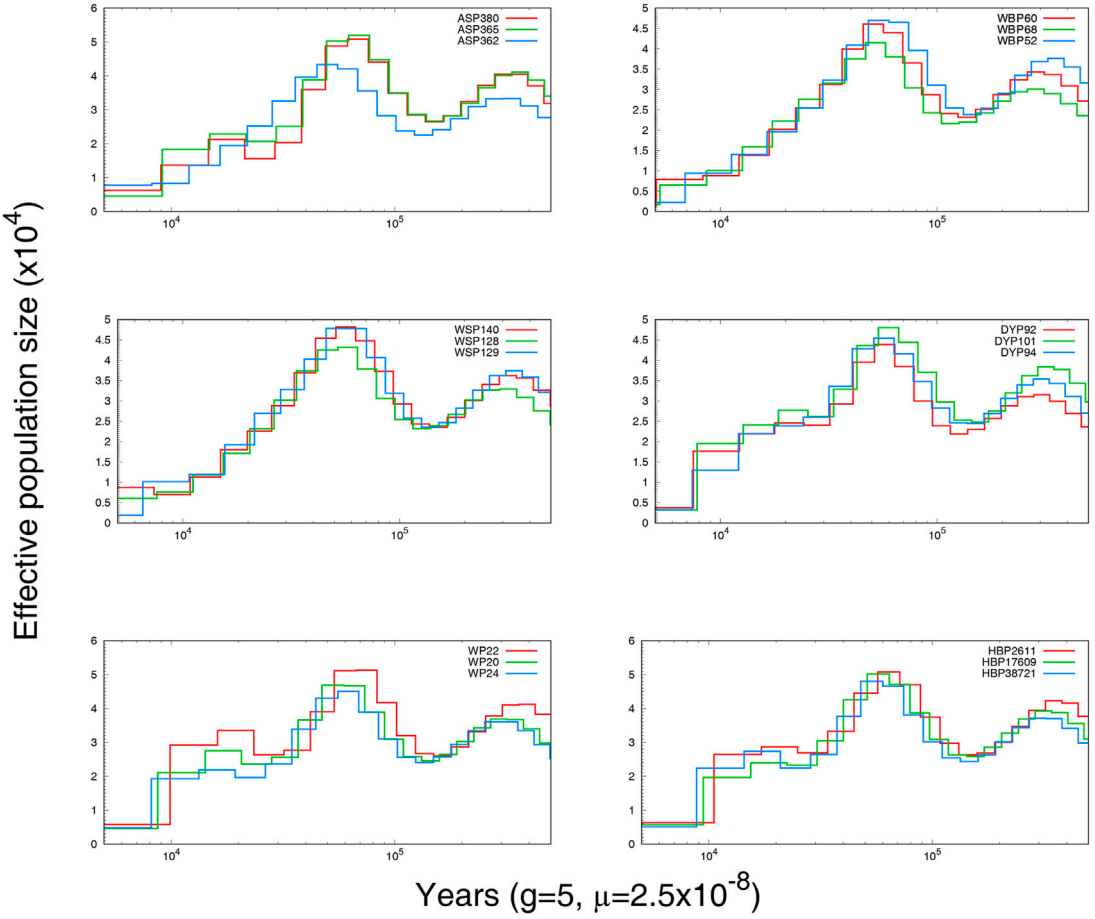


FIGURE 3
The effective population size and history of Anhui pig.

INSL6, MAP3K12, and MAPK1) and lipid- and meat-related processes (SNX19 and MSTN).

3.3.2 The selective regions of Anhui pig populations compared to those of the commercial pig population

To capture potential genes that underwent divergent selection between the Anhui and commercial pig populations, we estimated the genome-wide values of $\log_2$ ($\theta\pi$ ratio) and FST based on the whole-genome resequencing data of 177 samples from ten pig populations. The genome distribution of the two statistics were shown in Figures 5A,B. Using the top 5% FST cutoff (FST ≥ 0.2527) and $\log_2$ ($\theta\pi$ ratio) cutoff [$\log_2$ ($\theta\pi$ ratio) ≥ 0.0011], we detected nine hundred and sixty-seven selective regions containing 331 candidate genes (38.7 Mb of the genome; Figure 5C and Supplementary Tables S9, S10). All 331 genes were used for GO and KEGG analyses (Supplementary Figures S3, S4; Supplementary Tables S11, S12). Functional enrichment analyses revealed that the selected genes may play an important role in reproduction (*CABS1*, *MAPK1*, *IGF1R*, *INSR*, *LIMK2*, and *PATZ1*), meat quality and fat deposition capacity (*MC5R*, *MSTN*, *PRKG1*, *CREBBP*, and *ADCY9*), and ear size (*MSRB3* and *SOX5*).

4 Discussion

The pig, as one of the first domesticated animals, plays an important role in many aspects of human life. In recent years, the continued protection and utilization of the biological seed industry has attracted significant attention in China. The successively implemented “The third national census of livestock and poultry genetic resources” and “Precise identification of livestock and poultry germplasm resources” could help us to better understand the differences among pig breeds, and to protect and utilize the unique Chinese pig germplasm. To systematically analyze the genetic relationships, genetic diversity, population structure, and

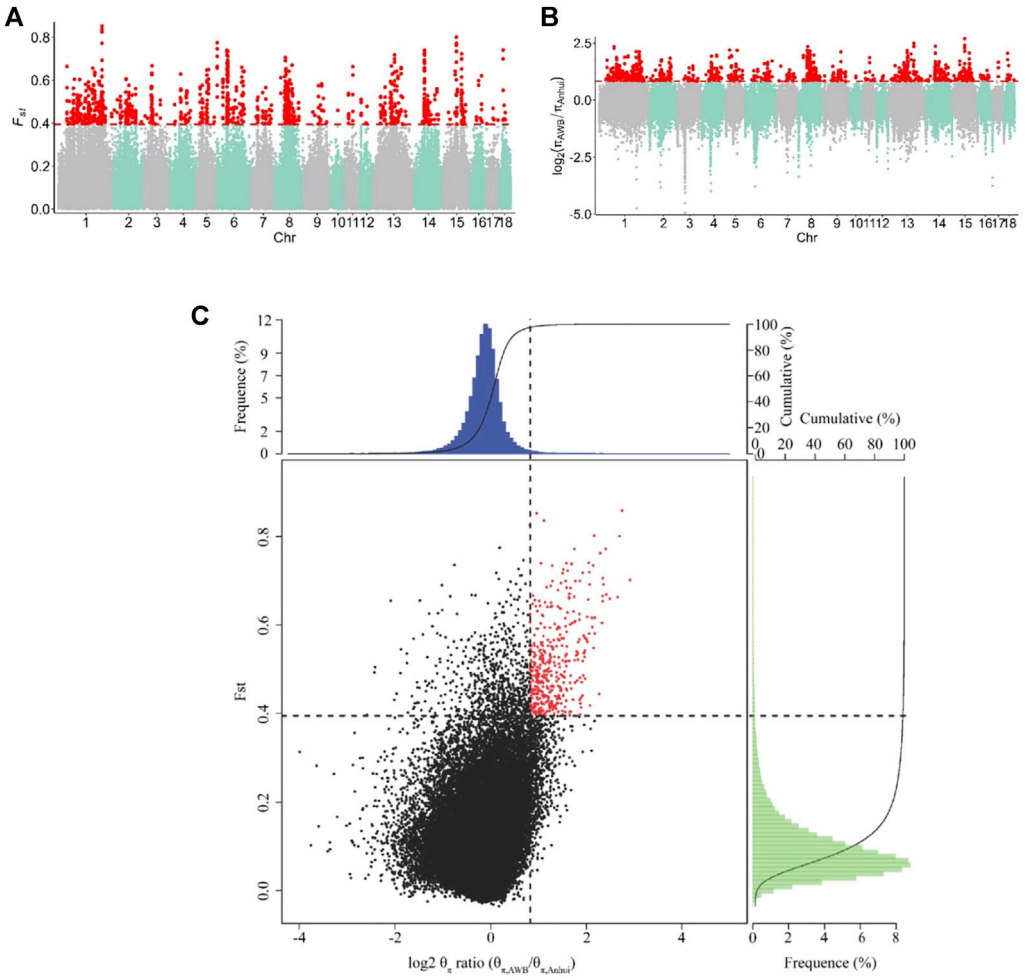


FIGURE 4
Identification of genomic regions with strong selective sweep signals in Anhui six pig population, which are calculated in a 40-kb sliding window approach with 20 kb step-size. **(A)** Distribution of FST values among autosome chromosomes. The red line represents the 0.01 level. **(B)** Distribution of $\log_2$ (θπ ratio) among autosomal chromosomes. The red line represents the 0.01 level. **(C)** The final selection regions based on two statistics. Points located to the right of the vertical dashed lines (corresponding to 1% right tails of the empirical $\log_2$ (θπ ratio) distribution, where $\log_2$ (θπ ratio) is 0.824992) and above the horizontal dashed line (1% above tail of the empirical FST distribution, where FST is 0.395220) were identified as selected regions for Anhui pig population (red points). The AWB refer to Asia wild boar, the Anhui refer to Anhui pig population.

selection signatures resulting from natural and artificial selection in the Anhui pig populations, 150 unrelated samples were collected from six pig breeds along with sequencing data from 40 foreign pigs. In this study, the genetic variation in the Anhui pig populations was surveyed, using blood samples from pig conservation farms, from which DNA was isolated and sequenced in order to construct a genetic conservation system. Going forward, this information can serve to guide the protection and utilization of indigenous pig breeds in Anhui. Moreover, the genetic structure of this population has also been elucidated. The WSP and WBP descended from the same ancestor, and the other four breeds from another ancestor. We also found that the Anhui pig populations exhibited a highly similar pattern of historical fluctuation in effective population size at approximately 10,000 BP, which is likely the consequence of Neolithic domestication events. In addition, several genes have been found to be associated with important economic traits.

4.1 The selected genes of Anhui pig populations compared to those of the Asian wild boar population

Domestic pigs have higher fertility rates than those of wild boars. GO analysis revealed several genes related to reproduction, including those related to reproductive process (GO: 0022414, five genes), reproduction (GO: 0000003, six genes), and fertilization (GO: 0007283, two genes). The acrosome

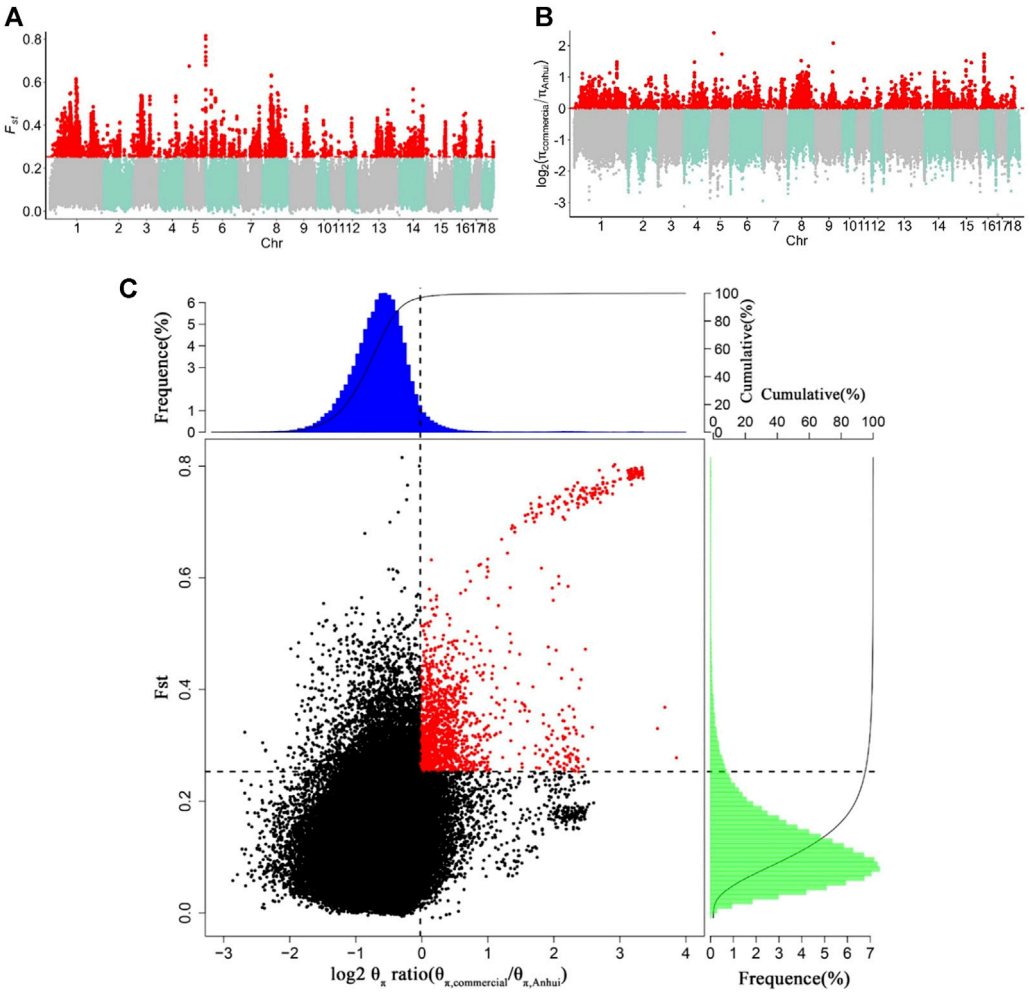


FIGURE 5
Identification of genomic regions with strong selective sweep signals between Anhui six pig population and four commercial pig population, which are calculated in a 40-kb sliding window approach with 20 kb step-size. **(A)** Distribution of F_{ST} values among autosome chromosomes. The red line represents the 0.05 level. **(B)** Distribution of $\log_2$ (θπ ratio) among autosomal chromosomes. The red line represents the 0.05 level. **(C)** The final selection regions based on two statistics. Points located to the right of the vertical dashed lines (corresponding to 5% right tails of the empirical $\log_2$ (θπ ratio) distribution, where $\log_2$ (θπ ratio) is 0.0.0011) and above the horizontal dashed line (5% above tail of the empirical FST distribution, where FST is 0.2527) were identified as selected regions for Anhui pig population (red points). The commercial refers to four commercial pigs, the Anhui refer to Anhui pig population.

reactions decreased significantly after the porcine capacitated sperm were treated with anti-pCABS1 antiserum, suggesting that porcine expression of CABS1 plays an important role in acrosome reactions (Shawki et al., 2016). In this study, two variants of the CABS1 gene were identified (Supplementary Table S13; exon1: c; G1047T: p.S349S; exon1:c.C1069G: p.L357V). Notably, the "exon1:c.C1069G: p.L357V" variant results in a Leu to Val substitution, which may in turn alter the expression of CABS1. The CABS1 gene was identified as a selected gene in ASP in our previous study (Zhang et al., 2020), which suggests that CABS1 is commonly selected within the Anhui pig populations. Another gene involved in reproduction, insulin-like 6 (INSL6), is a member of the insulin superfamily that plays an important role in the progression of spermatogenesis. Moreover, a deficiency in INSL6 can result in varying levels of male infertility (Lok et al., 2000; Brailoiu et al., 2005; Ivell and Grutzner. 2009; Chan et al., 2011). In this study, a missense variant of INSL6 (Supplementary Table S14; exon2: c.C403G: p.Q135E) was found among the Anhui pig breeds. This variation promotes a Gln-to-Glu substitution, implying that variations may impact the progression of spermatogenesis. Mitogen-activated protein kinase 12 (MAP3K12) is involved in the MAPK signaling pathway and is specifically selected in a high fecundity goat lineage, indicating its role in reproduction

(Lai et al., 2016). MAPK1 plays an important role in embryonic and placental development (Hatano et al., 2003). Furthermore, a signal adequate for promoting trophoblast proliferation and invasion can decrease below thresholds if MAPK1 protein expression is insufficient (Saba et al., 2003; Jeong et al., 2013). In a previous study, population high-density SNP array analysis of multiparous and uniparous sheep revealed that MAPK1 is a selected gene among Kazakhstan sheep (Wang et al., 2020), highlighting its role in reproduction. One synonymous SNV was also identified in MAPK1 (Supplementary Table S15; exon1:c.C144T: p.L48L).

Genes related to lipid and meat traits have also been identified. Sorting nexin 19 (SNX19) is primarily associated with meat quality and plays a vital role in the fat subnetwork (Kogelman et al., 2014). Interestingly, SNX19 was identified as a selected gene in a previous pig study (Guo et al., 2021). Five nonsynonymous variants were identified for SNX19 (Supplementary Table S16; exon11: c. A1037G: p.N346S, exon7:c.C487T: p.P163S, exon4: c. G94T: p.V32L, exon2: c.C1834T: p.L612F, exon1: c. G37C: p.A13P), and the variants with observable phenotypes were used to discriminate the functional site. We also identified myostatin (MSTN), a transcription factor belonging to the transforming growth factor beta (TGFβ) superfamily which plays an important role in muscle and fat development in pigs. Gene ontology annotation showed that among the Anhui pig populations, MSTN was enriched for growth (GO: 0040007), growth factor activity (GO: 0008083), and protein binding (GO: 0003824). Furthermore, a study by Deng et al. (2012) revealed that MSTN impacts the development of the longissimus muscle and the rectus superior muscle. Transgenic pigs with MSTN displayed decreased muscle growth and significant increases in intramuscular fat (Yang et al., 2009). Additionally, it has been shown that MSTN has the potential to regulate adipogenesis of mesenchymal stem cells during the determination and differentiation phases (Sun et al., 2016). MSTN inhibits adipogenesis in preadipocytes, (Hirai et al., 2007; Carrarelli et al., 2015), whereas it promotes adipogenesis in pluripotent stem cells (Feldman et al., 2006; Pantoja et al., 2008). In animals, MSTN-knockout typically reduces fat mass and resistance to diet-induced obesity (Guo et al., 2009; Gu et al., 2016).

4.2 The selected genes of Anhui pig populations compared to those of the commercial pig population

We identified several genes related to reproduction among the GO terms and KEGG pathways, with each playing an important role in spermatogenesis, ovarian steroidogenesis, the estrogen signaling pathway, the FoxO signaling pathway, or oocyte meiosis. Of these genes, calcium-binding protein spermatid associated 1(CABS1) and mitogen-activated protein kinase 1 (MAPK1) were those found in the selective regions of the Anhui pig population, but not in the selective regions of the Asian wild boar. The insulin-like growth factor 1 receptor (IGF1R) is widely expressed in mammals (including pigs, sheep, goats, and ducks) and is involved in growth, carcass traits, and reproductive performance (Liu et al., 1993; Terman, 2011; Hoopes et al., 2012; Kijas et al., 2012; Zhao et al., 2018; Jin et al., 2020; Grochowska et al., 2021; Jiang et al., 2021). In female mice, IGF1R was found to play a vital role in steroidogenesis, follicle survival, and fertility (Baumgarten et al., 2017), where it helped transfer nutrients to the fetus (Hellström et al., 2016). Zhou et al. (2021) found that IGF1R is also necessary for epithelial differentiation and normal uterine preparation for embryo implantation (Zhou et al., 2021). Given this background, IGF1R likely contributes to phenotypic differences in growth and reproduction among pigs. In this study, 450 variations were identified in the IGF1R gene (Supplementary Table S18), and only seven were located in exons, none of which resulted in amino acid changes. Conversely, three variations were found within the 3′UTR, and 30 within the 5′UTR that may regulate the function of the IGF1R. Insulin receptor (INSR), a tyrosine kinase receptor, affects development and growth (Hubbard. 2013; Bedinger and Adams. 2015). INSR has been validated as a polycystic ovary syndrome risk locus using a very large, well-designed case-control GWAS (Shi et al., 2012). Meanwhile, INSR has been confirmed to be necessary for optimal endometrial proliferation and implantation (Sekulovski et al., 2021). Furthermore, INSR plays a crucial role in controlling adipose tissue development and adipocyte survival (Boucher et al., 2012; Cignarelli et al., 2019). INSR may be responsible for fat accumulation in adipose tissues, which could induce angiopoietin-like 8, inhibit lipolysis, control postprandial fat storage in white adipose tissue, and direct that fatty acids be stored in the adipose tissue during the fed state (Ahbara et al., 2019). The testis-specific isoform of LIM kinase 2 (tLIMK2) plays a key role in the progression of spermatogenesis (Takahashi et al., 2002). The weight of LIMK2−/− mice sharply decreased by 20% compared to that of control mice. Furthermore, inhibition of LIMK1/2 activity results in failure of embryo cleavage and blastocyst formation (Duan et al., 2018). POZ/BTB and AT hook containing zinc finger 1(PATZ1), a zinc finger protein, can affect the basal activity of different promoters. PATZ1 plays a crucial role in embryonic development (Valentino et al., 2013) and normal male gametogenesis and impairment of PATZ1 result in disruption of testis cytoarchitecture and block spermatogenesis (Fedele et al., 2008).

According to previous research, indigenous Anhui pig breeds exhibit significant differences in meat quality traits and fat deposition capacity compared to commercial pig breeds (Xu et al., 2018; Hu et al., 2019; Li et al., 2021). In the present study, several genes were found to be associated with fat deposition and other carcass quality traits. The melanocortin five receptor gene (MC5R), a member of the G protein-coupled receptor

superfamily, participates in lipid production, fatty acid oxidation of skeletal muscle, and lipolysis of dipocytes. MC5R is associated with the metabolism of skeletal muscle, fatty acids, and fat cells, as well as dysfunction of the exocrine gland under abnormal expression (Zhang et al., 2011). The polymorphisms of MC5R are associated with obesity in humans (Chagnon et al., 1997; Valli et al., 2008). Furthermore, MC5R deficiency accelerates lipolysis and reduces oil secretion (Chen et al., 1997). Whole genome analysis of 46 cattle from six representative Chinese breeds together with international breeds has shown that MC5R was under-selected (Mei et al., 2018). In the pig genome, MC5R was found close to the 97.625–98.725 Mb regions on swine chromosome 6 (SSC6). The regions of SSC6 are associated with lipid metabolism, back fat thickness, and intramuscular fat percentage (Switonski et al., 2013; Ma et al., 2018), and play a vital role in proinflammatory activity (Jun et al., 2010). Among the pigs in this study, three variations were found in the 3′UTR of the MC5R gene, and four variations were located in the exon (Supplementary Table S19). Two exonic variants (exon3: c.A952G:p. S318G; exon3:c.A580G:p.T194A) resulted in amino acid changes from Ser to Gly and Thr to Ala. The exon1: c. G92A created a premature stop codon in the MC5R amino acid sequence. Additionally, kinase cGMP-dependent 1 (PRKG1) has been shown to regulate adipocytes lipolysis and plays an important role in pig fatty acid composition (Revilla et al., 2017). Interestingly, PRKG1 exhibits differential expression between the high and low fatty acid composition groups in muscle according to RNA-Seq analysis (Puig et al., 2014). Moreover, PRKG1 knockout mice have decreased triglyceride stores in brown adipose tissue (Amieux and Mcknight, 2010). The taste and quality of cooked and cured meat are directly affected by the oxidative stability of the muscle, which is related to fatty acid (FA) composition (Serra et al., 1998; Wood et al., 2008). We also found two genes, CREB binding protein (CREBBP) and adenylate cyclase 9 (ADCY9), which are involved in cAMP signaling. cAMP signaling regulates energy homeostasis in multiple tissues by mediating the action of metabolism-controlling hormones such as glucagon and epinephrine (Ravnskjaer et al., 2016). A previous GWAS and post-GWAS study discovered that CREBBP and ADCY9 are located inside regions that are significantly associated with meat pH (Verardo et al., 2017). For ADCY9, there were 13 exonic variants, three of which were missense variants (Supplementary Table S20, exon8: c.A2594G:p.H865R; exon8:c.T2669C:p.M890T; exon11: c.G3943T:p.G1315C), results in amino acid changes that may impact meat quality. CREBBP regulates a plethora of metabolic target genes involved in glucose metabolism (Ravnskjaer et al., 2016), which can also potentially affect meat quality due to acidification. In this study, one missense variant was identified (Supplementary Table S21, exon22:c.A3907G:p. I1303V) in which the amino acid Ile was changed to Val.

Two selected genes were found to be associated with ear size phenotypic traits. The size and type of the ears are important conformational characteristics that distinguish pig breeds. Many indigenous Chinese pig breeds, such as the six Anhui pig populations, have unusually large floppy ears. In contrast, commercial breeds (Duroc, Landrace, Yorkshire, and Pitelan) have smaller, more erect ears. In this study, we identified two important candidate genes influenced by selection. The first is methionine sulfoxide reductase B3 (MSRB3), which has been shown to be associated with ear size in sheep, goats, dogs, and pigs (Webster et al., 2015; Zhang et al., 2015; Chen et al., 2018; Kumar et al., 2018; Zhao et al., 2020; Posbergh and Huson. 2021). In the present study, a variant in the 3′-UTR of MSRB3 was identified among the Anhui pig breeds (Supplementary Table S17; g. T29862412C), suggesting that it could be the source of their erect ear phenotype. SRY-box transcription factor 5 (SOX5) encodes a member of the SOX transcription factor family, and has been shown to play a role in chondrogenesis (Lefebvre et al., 1998; Smits et al., 2001). Additionally, it has been identified as a selected gene in Duroc pigs and Duroc × Korean native pigs, in which it plays a key role in ear morphology (Edea et al., 2017). Moreover, mutations in genes encoding homeobox transcription factors are often responsible for the defective development of the outer ear (Fekete, 1999). Positive selection of MSRB3 and SOX5 among the Anhui pig populations, but not among the commercial pig breeds, may be the genetic mechanism behind the ear size phenotype.

Although the present study has produced many interesting findings, it has its limitations. Firstly, as the phenotypic values were not collected, functional experimental assays are still needed to further validate the associations between the phenotypes and genotypes of the mentioned variants and to identify the targets involved in reproduction, lipid and meat quality, and ear size.

In this study, we generated a novel catalog of population genomic data on the Anhui pig populations using whole-genome resequencing. Population genomic analyses have further elucidated genomic variation, population structure, demographic history, and signatures of selection among Anhui pigs. Moreover, we discovered several potential signatures of selection associated with the domestication characteristics of the Anhui pig populations, with selected regions involved in reproduction, lipid and meat characteristics, and ear size. These findings substantially expand the catalogue of genetic variants among pigs, and the newly generated genome-wide data are a valuable resource for future genetic studies and those to improve genome-assisted breeding of pigs and other domestic animals.

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 the recommendations of the Animal Care Committee of Anhui Academy of Agricultural Sciences (Hefei, People's Republic of China). The protocol was approved by the Animal Care Committee of Anhui Academy of Agricultural Sciences (No. AAAS2020-04).

Author contributions

Conceptualization: WZ and CW; Methodology: WZ and YJ; Investigation: WZ and XiL; Resources: XiL, YJ, SS, CX, and XuL; Data curation: WZ; Writing—original draft preparation: WZ; Funding acquisition: WZ, MZ, and CW.

Funding

This work was supported by the grants from Anhui Academy of Agricultural Sciences Key Laboratory Project (No. 2021YL023), Anhui Province Financial Fund for Modern Seed Industry Project, Anhui Province Natural Science Foundation Youth Fund Project (2108085QC135); the Special Fund for Anhui Agricultural Research System (AHCYJSTX-05-12, AHCYJSTX-05-23); the Anhui Provincial Key Laboratory of Livestock and Poultry Product Safety Engineering Young Talents Support Engineering Innovation Guidance Fund (XMT 2022-09).

Acknowledgments

We thank many people not listed as authors who provided help, encouragement and feedback.

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.2022.1022261/full#supplementary-material

References

Ahbara, A., Bahbahani, H., Almathen, F., Al Abri, M., Agoub, M. O., Abeba, A., et al. (2019). Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep. *Front. Genet.* 9, 699. doi:10.3389/fgene.2018.00699

Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., et al. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47, 217–225. doi:10.1038/ng.3199

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109

Amieux, P. S., and McKnight, G. S. (2010). Cyclic nucleotides converge on Brown adipose tissue differentiation. *Sci. Signal.* 3, pe2. doi:10.1126/scisignal.3104pe2

Baumgarten, S. C., Armouti, M., Ko, C., and Stocco, C. (2017). IGF1R expression in ovarian granulosa cells is essential for steroidogenesis, follicle survival, and fertility in female mice. *Endocrinology* 158, 2309–2318. doi:10.1210/en.2017-00146

Bedinger, D. H., and Adams, S. H. (2015). Metabolic, anabolic, and mitogenic insulin responses: A tissue-specific perspective for insulin receptor activators. *Mol. Cell. Endocrinol.* 415, 143–156. doi:10.1016/j.mce.2015.08.013

Boitard, S., Rodriguez, W., Jay, F., Mona, S., and Austerlitz, F. (2016). Inferring population size history from large samples of genome-wide molecular data-an approximate Bayesian computation approach. *PLoS Genet.* 12, e1005877. doi:10.1371/journal.pgen.1005877

Boucher, J., Mori, M. A., Lee, K. Y., Smyth, G., Liew, C. W., Macotela, Y., et al. (2012). Impaired thermogenesis and adipose tissue development in mice with fat-specific disruption of insulin and IGF-1 signalling. *Nat. Commun.* 3, 902. doi:10.1038/ncomms1905

Brailoiu, G. C., Dun, S. L., Yin, D., Yang, J., Chang, J. K., and Dun, N. J. (2005). Insulin-like 6 immunoreactivity in the mouse brain and testis. *Brain Res.* 1040, 187–190. doi:10.1016/j.brainres.2005.01.077

Carrarelli, P., Yen, C. F., Arcuri, F., Funghi, L., Tosti, C., Wang, T. H., et al. (2015). Myostatin, follistatin and activin type II receptors are highly expressed in adenomyosis. *Fertil. Steril.* 104, 744–752. doi:10.1016/j.fertnstert.2015.05.032

Chagnon, Y. C., Chen, W. J., Pérusse, L., Chagnon, M., Nadeau, A., Wilkison, W. O., et al. (1997). Linkage and association studies between the melanocortin receptors 4 and 5 genes and obesity-related phenotypes in the Québec Family Study. *Mol. Med.* 3, 663–673. doi:10.1007/bf03401705

Chan, L. J., Hossain, M. A., Samuel, C. S., Separovic, F., and Wade, J. D. (2011). The relaxin peptide family-structure, function and clinical applications. *Protein Pept. Lett.* 18, 220–229. doi:10.2174/092986611794578396

Chen, C., Liu, C., Xiong, X., Fang, S., Yang, H., Zhang, Z., et al. (2018). Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Sel. Evol.* 50, 72. doi:10.1186/s12711-018-0442-6

Chen, W., Kelly, M. A., Opitz-Araya, X., Thomas, R. E., Low, M. J., and Cone, R. D. (1997). Exocrine gland dysfunction in MC5-R-deficient mice: Evidence for coordinated regulation of exocrine gland function by melanocortin peptides. *Cell.* 91, 789–798. doi:10.1016/s0092-8674(00)80467-5

China National Commission of Animal Genetic Resource (2011). *Animal genetic resource in China. Pigs*. Beijing, China: China Agriculture Press.

Cignarelli, A., Genchi, V. A., Perrini, S., Natalicchio, A., Laviola, L., and Giorgino, F. (2019). Insulin and insulin receptors in adipose tissue development. *Int. J. Mol. Sci.* 20, 759. doi:10.3390/ijms20030759

Deng, B., Wen, J., Ding, Y., Gao, Q., Huang, H., Ran, Z., et al. (2012). Functional analysis of pig myostatin gene promoter with some adipogenesis- and myogenesis-related factors. *Mol. Cell. Biochem.* 363, 291–299. doi:10.1007/s11010-011-1181-y

Duan, X., Zhang, H. L., Wu, L. L., Liu, M. Y., Pan, M. H., Ou, X. H., et al. (2018). Involvement of LIMK1/2 in actin assembly during mouse embryo development. *Cell. Cycle* 17, 1381–1389. doi:10.1080/15384101.2018.1482138

Edea, Z., Hong, J. K., Jung, J. H., Kim, D. W., Kim, Y. M., Kim, E. S., et al. (2017). Detecting selection signatures between Duroc and Duroc synthetic pig populations using high-density SNP chip. *Anim. Genet.* 48, 473–477. doi:10.1111/age.12559

Fang, M., Larson, G., Ribeiro, H. S., Li, N., and Andersson, L. (2009). Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* 5, e1000341. doi:10.1371/journal.pgen.1000341

Fedele, M., Franco, R., Salvatore, G., Paronetto, M. P., Barbagallo, F., Pero, R., et al. (2008). PATZ1 gene has a critical role in the spermatogenesis and testicular tumours. *J. Pathol.* 215, 39–47. doi:10.1002/path.2323

Fekete, D. M. (1999). Development of the vertebrate ear: Insights from knockouts and mutants. *Trends Neurosci.* 22, 263–269. doi:10.1016/s0166-2236(98)01366-6

Feldman, B. J., Streeper, R. S., Farese, R. V., Jr, and Yamamoto, K. R. (2006). Myostatin modulates adipogenesis to generate adipocytes with favorable metabolic effects. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15675–15680. doi:10.1073/pnas.0607501103

Frantz, L. A., Schraiber, J. G., Madsen, O., Megens, H. J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148. doi:10.1038/ng.3394

Grochowska, E., Lisiak, D., Akram, M. Z., Adeniyi, O. O., Lühken, G., and Borys, B. (2021). Association of a polymorphism in exon 3 of the IGF1R gene with growth, body size, slaughter and meat quality traits in Colored Polish Merino sheep. *Meat Sci.* 172, 108314. doi:10.1016/j.meatsci.2020.108314

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi:10.1038/nature11622

Gu, H., Cao, Y., Qiu, B., Zhou, Z., Deng, R., Chen, Z., et al. (2016). Establishment and phenotypic analysis of an Mstn knockout rat. *Biochem. Biophys. Res. Commun.* 477, 115–122. doi:10.1016/j.bbrc.2016.06.030

Guo, L., Sun, H., Zhao, Q., Xu, Z., Zhang, Z., Liu, D., et al. (2021). Positive selection signatures in Anqing six-end-white pig population based on reduced-representation genome sequencing data. *Anim. Genet.* 52, 143–154. doi:10.1111/age.13034

Guo, T., Jou, W., Chanturiya, T., Portas, J., Gavrilova, O., and McPherron, A. C. (2009). Myostatin inhibition in muscle, but not adipose tissue, decreases fat mass and improves insulin sensitivity. *PLoS One* 4, e4937. doi:10.1371/journal.pone.0004937

Hatano, N., Mori, Y., Oh-hora, M., Kosugi, A., Fujikawa, T., Nakai, N., et al. (2003). Essential role for ERK2 mitogen-activated protein kinase in placental development. *Genes. cells.* 8, 847–856. doi:10.1046/j.1365-2443.2003.00680.x

Hellström, A., Ley, D., Hansen-Pupp, I., Hallberg, B., Ramenghi, L. A., Löfqvist, C., et al. (2016). Role of insulinlike growth factor 1 in fetal development and in the early postnatal life of premature infants. *Am. J. Perinatol.* 33, 1067–1071. doi:10.1055/s-0036-1586109

Hirai, S., Matsumoto, H., Hino, N., Kawachi, H., Matsui, T., and Yano, H. (2007). Myostatin inhibits differentiation of bovine preadipocyte. *Domest. Anim. Endocrinol.* 32, 1–14. doi:10.1016/j.domaniend.2005.12.001

Hoopes, B. C., Rimbault, M., Liebers, D., Ostrander, E. A., and Sutter, N. B. (2012). The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* 23, 780–790. doi:10.1007/s00335-012-9417-z

Hu, H., Wu, C. D., Ding, Y. Y., Zhang, X. D., Yang, M., Wen, A. Y., et al. (2019). Comparative analysis of meat sensory quality, antioxidant status, growth hormone and orexin between Anqingliubai and Yorkshire pigs. *J. Appl. Anim. Res.* 47, 357–361. doi:10.1080/09712119.2019.1643729

Hubbard, S. R. (2013). The insulin receptor: Both a prototypical and atypical receptor tyrosine kinase. *Cold Spring Harb. Perspect. Biol.* 5 (3), a008946. doi:10.1101/cshperspect.a008946

Ivell, R., and Grutzner, F. (2009). Evolution and male fertility: Lessons from the insulin-like factor 6 gene (Insl6). *Endocrinology* 150, 3986–3990. doi:10.1210/en.2009-0691

Jeong, W., Kim, J., Bazer, F. W., and Song, G. (2013). Epidermal growth factor stimulates proliferation and migration of porcine trophectoderm cells through protooncogenic protein kinase 1 and extracellular-signal-regulated kinases 1/2 mitogen-activated protein kinase signal transduction cascades during early pregnancy. *Mol. Cell. Endocrinol.* 81, 302–311. doi:10.1016/j.mce.2013.08.024

Jiang, F., Lin, R., Xiao, C., Xie, T., Jiang, Y., Chen, J., et al. (2021). Analysis of whole-genome re-sequencing data of ducks reveals a diverse demographic history and extensive gene flow between Southeast/South Asian and Chinese populations. *Genet. Sel. Evol.* 53, 35. doi:10.1186/s12711-021-00627-0

Jin, M., Lu, J., Fei, X., Lu, Z., Quan, K., Liu, Y., et al. (2020). *Genetic Signatures of Selection for Cashmere Traits in Chinese Goats*, 10.*Anim. (Basel)*

Jun, D. J., Na, K. Y., Kim, W., Kwak, D., Kwon, E. J., Yoon, J. H., et al. (2010). Melanocortins induce interleukin 6 gene expression and secretion through melanocortin receptors 2 and 5 in 3T3-L1 adipocytes. *J. Mol. Endocrinol.* 44, 225–236. doi:10.1677/JME-09-0161

Kerstens, H. H., Kollers, S., Kommadath, A., Del Rosario, M., Dibbits, B., Kinders, S. M., et al. (2009). Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC genomics* 10, 4. doi:10.1186/1471-2164-10-4

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10, e1001258. doi:10.1371/journal.pbio.1001258

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., et al. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinforma.* 12, 231. doi:10.1186/1471-2105-12-231

Kogelman, L. J., Cirera, S., Zhernakova, D. V., Fredholm, M., Franke, L., and Kadarmideen, H. N. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA sequencing in a porcine model. *BMC Med. Genomics* 7, 57. doi:10.1186/1755-8794-7-57

Kumar, C., Song, S., Dewani, P., Kumar, M., Parkash, O., Ma, Y., et al. (2018). Population structure, genetic diversity and selection signatures within seven indigenous Pakistani goat populations. *Anim. Genet.* 49, 592–604. doi:10.1111/age.12722

Kuo, I. Y., and Ehrlich, B. E. (2015). Signaling in muscle contraction. *Cold Spring Harb. Perspect. Biol.* 7, a006023. doi:10.1101/cshperspect.a006023

Lai, F. N., Zhai, H. L., Cheng, M., Ma, J. Y., Cheng, S. F., Ge, W., et al. (2016). Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (*Capra hircus*). *Sci. Rep.* 6, 38096. doi:10.1038/srep38096

Lefebvre, V., Li, P., and de Crombrugghe, B. (1998). A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.* 17, 5718–5733. doi:10.1093/emboj/17.19.5718

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

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole genome sequences. *Nature* 475, 493–496. doi:10.1038/nature10231

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, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi:10.1093/bioinformatics/btr509

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi:10.1038/ng.2811

Li, X. J., Liu, L. Q., Dong, H., Yang, J. J., Wang, W. W., Zhang, Q., et al. (2021). Comparative genome-wide methylation analysis of longissimus dorsi muscles in Yorkshire and Wannanhua pigs. *Anim. Genet.* 52, 78–89. doi:10.1111/age.13029

Liu, J. P., Baker, J., Perkins, A. S., Robertson, E. J., and Efstratiadis, A. (1993). Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell.* 75, 59–72. doi:10.1016/s0092-8674(05)80084-4

Lok, S., Johnston, D. S., Conklin, D., Lofton-Day, C. E., Adams, R. L., Jelmberg, A. C., et al. (2000). Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat. *Biol. Reprod.* 62, 1593–1599. doi:10.1095/biolreprod62.6.1593

Ma, Y., Zhang, S., Zhang, K., Fang, C., Xie, S., Du, X., et al. (2018). *Genomic Analysis to Identify Signatures of Artificial Selection and Loci Associated with Important Economic Traits in Duroc Pigs*, 8. Bethesda, Md: G3, 3617–3625.

Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., et al. (2018). Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol. Biol. Evol.* 35, 688–699. doi:10.1093/molbev/msx322

Nei, M., and Li, W-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5269–5273. doi:10.1073/pnas.76.10.5269

Pantoja, C., Huff, J. T., and Yamamoto, K. R. (2008). Glucocorticoid signaling defines a novel commitment state during adipogenesis *in vitro*. *Mol. Biol. Cell.* 19, 4032–4041. doi:10.1091/mbc.e08-04-0420

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31, 1929–1936. doi:10.1093/molbev/msu136

Posbergh, C. J., and Huson, H. J. (2021). All sheeps and sizes: A genetic investigation of mature body size across sheep breeds reveals a polygenic nature. *Anim. Genet.* 52, 99–107. doi:10.1111/age.13016

Puig, A., Ramayo-Caldas, Y., Corominas, J., Estelle, J., Perez-Montarelo, D., Hudson, N. J., et al. (2014). Differences in muscle Transcriptome among pigs phenotypically extreme for fatty acid composition. *PLoS One* 9, e99720. doi:10.1371/journal.pone.0099720

Ravnskjaer, K., Madiraju, A., and Montminy, M. (2016). Role of the cAMP pathway in glucose and lipid metabolism. *Handb. Exp. Pharmacol.* 233, 29–49. doi:10.1007/164_2015_32

Revilla, M., Puig-Oliveras, A., Castelló, A., Crespo-Piazuelo, D., Paludo, E., Fernández, A. I., et al. (2017). A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. *PLoS One* 12, e0177014. doi:10.1371/journal.pone.0177014

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

Saba, M. K., Vella, F. D., Vernay, B., Voisin, L., Chen, L., Labrecque, N., et al. (2003). An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep.* 4, 964–968. doi:10.1038/sj.embor.embor939

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning: A laboratory manual*. 3rd ed. New York: Cold Spring Harbor Laboratory Press.

Sekulovski, N., Whorton, A. E., Shi, M., Hayashi, K., and MacLean, J. A. (2021). Insulin signaling is an essential regulator of endometrial proliferation and implantation in mice. *FASEB J.* 35, e21440. doi:10.1096/fj.202002448R

Serra, X., Gil, C., Pérez-Enciso, M., Oliver, M., Vázquez, J., Gispert, M., et al. (1998). A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. *Livest. Prod. Sci.* 56, 215–223. doi:10.1016/s0301-6226(98)00151-1

Shawki, H. H., Kigoshi, T., Katoh, Y., Matsuda, M., Ugboma, C. M., Takahashi, S., et al. (2016). Identification, localization, and functional analysis of the homologues of mouse CABS1 protein in porcine testis. *Exp. Anim.* 65, 253–265. doi:10.1538/expanim.15-0104

Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z., et al. (2012). Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat. Genet.* 44, 1020–1025. doi:10.1038/ng.2384

Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J. M., Behringer, R. R., et al. (2001). The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell.* 1, 277–290. doi:10.1016/s1534-5807(01)00003-x

Sun, W. X., Dodson, M. V., Jiang, Z. H., Yu, S. G., Chu, W. W., and Chen, J. (2016). Myostatin inhibits porcine intramuscular preadipocyte differentiation *in vitro*. *Domest. Anim. Endocrinol.* 55, 25–31. doi:10.1016/j.domaniend.2015.10.005

Switonski, M., Mankowska, M., and Salamon, S. (2013). Family of melanocortin receptor (MCR) genes in mammals-mutations, polymorphisms and phenotypic effects. *J. Appl. Genet.* 54, 461–472. doi:10.1007/s13353-013-0163-z

Takahashi, H., Koshimizu, U., Miyazaki, J., and Nakamura, T. (2002). Impaired spermatogenic ability of testicular germ cells in mice deficient in the LIM-kinase 2 gene. *Dev. Biol.* 241, 259–272. doi:10.1006/dbio.2001.0512

Terman, A. (2011). The IGF1R gene: A new marker for reproductive performance traits in sows? *Acta Agric. Scand. Sect. A - Animal Sci.* 61, 67–71. doi:10.1080/09064702.2011.570780

Valentino, T., Palmieri, D., Vitiello, M., Simeone, A., Palma, G., Arra, C., et al. (2013). Embryonic defects and growth alteration in mice with homozygous disruption of the Patz1 gene. *J. Cell. Physiol.* 228, 646–653. doi:10.1002/jcp.24174

Valli, K., Suviolahti, E., Schalin-Jäntti, C., Ripatti, S., Silander, K., Oksanen, L., et al. (2008). Further evidence for the role of ENPP1 in obesity: Association with morbid obesity in Finns. *Obes. (Silver Spring, Md.)* 16, 2113–2119. doi:10.1038/oby.2008.313

Verardo, L. L., Sevón-Aimonen, M. L., Serenius, T., Hietakangas, V., and Uimari, P. (2017). Whole-genome association analysis of pork meat pH revealed three significant regions and several potential genes in Finnish Yorkshire pigs. *BMC Genet.* 18, 13. doi:10.1186/s12863-017-0482-x

Wang, C., Wang, H., Zhang, Y., Tang, Z., Li, K., and Liu, B. (2015). Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol. Ecol. Resour.* 15, 414–424. doi:10.1111/1755-0998.12311

Wang, K., Li, M., and Hakonarson, H. (2011). Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wang, Y., Niu, Z., Zeng, Z., Jiang, Y., Jiang, Y., Ding, Y., et al. (2020). Using high-density SNP array to reveal selection signatures related to prolificacy in Chinese and Kazakhstan sheep breeds. *Animals.* 10, 1633. doi:10.3390/ani10091633

Webster, M. T., Kamgari, N., Perloski, M., Hoeppner, M. P., Axelsson, E., Hedhammar, Å., et al. (2015). Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC genomics* 16, 474. doi:10.1186/s12864-015-1702-2

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., et al. (2008). Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* 78, 343–358. doi:10.1016/j.meatsci.2007.07.019

Xu, J., Wang, C., Jin, E., Gu, Y., Li, S., and Li, Q. (2018). Identification of differentially expressed genes in longissimus dorsi muscle between Wei and Yorkshire pigs using RNA sequencing. *Genes. Genomics* 40, 413–421. doi:10.1007/s13258-017-0643-3

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). Gcta: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi:10.1016/j.ajhg.2010.11.011

Yang, J., Schulman, L., Pursel, V. G., Solomon, M. B., Mitchell, A. D., Zhao, B., et al. (2009). Expression of porcine myostatin prodomain genomic sequence leads to a decrease in muscle growth, but significant intramuscular fat accretion in transgenic pigs. *Transgenic Animal Res. Conf.* VIII, 15.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11597–11604. doi:10.1073/pnas.0801317105

Zhang, L., Li, W. H., Anthonavage, M., Pappas, A., Rossetti, D., Cavender, D., et al. (2011). Melanocortin-5 receptor and sebogenesis. *Eur. J. Pharmacol.* 660, 202–206. doi:10.1016/j.ejphar.2010.10.100

Zhang, W., Yang, M., Zhou, M., Wang, Y., Wu, X., Zhang, X., et al. (2020). Identification of signatures of selection by whole-genome resequencing of a Chinese native pig. *Front. Genet.* 11, 566255. doi:10.3389/fgene.2020.566255

Zhang, Y., Liang, J., Zhang, L., Wang, L., Liu, X., Yan, H., et al. (2015). Porcine methionine sulfoxide reductase B3: Molecular cloning, tissue-specific expression profiles, and polymorphisms associated with ear size in *Sus scrofa*. *J. Anim. Sci. Biotechnol.* 6, 60. doi:10.1186/s40104-015-0060-x

Zhao, F., Deng, T., Shi, L., Wang, W., Zhang, Q., Du, L., et al. (2020). Genomic scan for selection signature reveals fat deposition in Chinese indigenous sheep with extreme tail types. *Animals.* 10, 773. doi:10.3390/ani10050773

Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., et al. (2018). Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *GigaScience* 7, giy058. doi:10.1093/gigascience/giy058

Zhou, C., Lv, M., Wang, P., Guo, C., Ni, Z., Bao, H., et al. (2021). Sequential activation of uterine epithelial IGF1R by stromal IGF1 and embryonic IGF2 directs normal uterine preparation for embryo implantation. *J. Mol. Cell. Biol.* 13, 646–661. doi:10.1093/jmcb/mjab034

Zhu, J., Chen, C., Yang, B., Guo, Y., Ai, H., Ren, J., et al. (2015). A systems genetics study of swine illustrates mechanisms underlying human phenotypic traits. *BMC Genomics* 16, 88. doi:10.1186/s12864-015-1240-y

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fvets.2022.1009174

Check for updates

OPEN ACCESS

EDITED BY
Adnan Khan,
Agricultural Genomics Institute at Shenzhen (CAAS), China

REVIEWED BY
Saqib Umer,
University of Agriculture, Faisalabad, Pakistan
Ruiwen Fan,
Shanxi Agricultural University, China

*CORRESPONDENCE
Zhang Yun-hai
yunhaizhang@ahau.edu.cn

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

RECEIVED 01 August 2022
ACCEPTED 30 August 2022
PUBLISHED 03 November 2022

CITATION
Kai-yuan J, Yi-Wei Z, Ru-jun W, Khan IM and Yun-hai Z (2022) A genome-wide integrated analysis of lncRNA-mRNA in melanocytes from white and brown skin hair boer goats (*Capra aegagrus hircus*). *Front. Vet. Sci.* 9:1009174. doi: 10.3389/fvets.2022.1009174

COPYRIGHT
© 2022 Kai-yuan, Yi-Wei, Ru-jun, Khan and Yun-hai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A genome-wide integrated analysis of lncRNA-mRNA in melanocytes from white and brown skin hair boer goats (*Capra aegagrus hircus*)

Ji Kai-yuan[1,2], Zhao Yi-Wei[1], Wen Ru-jun[1], Ibrar Muhammad Khan[1,3] and Zhang Yun-hai[1,4]*

[1]Anhui Key Laboratory of Genetic Resources Protection and Biological Breeding for Livestock and Poultry, College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [2]Anhui Province Key Laboratory of Veterinary Pathobiology and Disease Control, College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [3]Anhui Province Key Laboratory of Embryo Development and Reproduction Regulation, Anhui Province Key Laboratory of Environmental Hormone and Reproduction, School of Biological and Food Engineering, Fuyang Normal University, Fuyang, China, [4]Linquan Comprehensive Experimental Station of Anhui Agricultural University, Anhui Agricultural University, Linquan, China

Long noncoding RNAs (lncRNAs) are involved in many biological processes and have been extensively researched. Nonetheless, literature focusing on the roles of lncRNA in melanocytes is limited. Melanocytes are located in the basal layer of the epidermis and determine the color of an animal's skin and hair by producing melanin. The mechanisms of melanogenesis remain unclear. Here, melanocytes from Boer goat skins were successfully isolated and verified using morphological observation, dopamine staining, silver ammonia staining, and immunohistochemical staining *in vitro*. Phenotypic testing revealed that melanocytes isolated from goat skins with white and brown hairs showed significant differences in proliferation, migration, and melanogenesis ($^{**}P < 0.01$). RNA sequencing was performed with the isolated melanocytes, and through bioinformatic analysis, several candidate lncRNAs and mRNAs involved in stage-specific melanogenesis were identified. Functional enrichment analysis indicated that miRNA precursors and *cis*-regulatory effects of lncRNAs were deeply dissected using the function prediction software. Multiple lncRNA–mRNA networks were presumed to be involved in melanocyte migration, proliferation, and melanogenesis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation. This research provided novel bioinformatic insights into the roles of lncRNAs in mammalian pigmentation.

KEYWORDS

lncRNA, melanocyte, *Capra hircus*, genome-wide analysis, mRNA

Introduction

Melanoblasts are found in the skin's basal layer and migrate to the epidermis to mature into melanocytes. Mature melanocytes play roles in the skin's innate immunity and determine skin color by producing melanin (1, 2). Melanosomes are unique organelles found in melanocytes that synthesize melanin granules under tyrosinase catalysis (1). Depending on the expressed regulatory genes, melanosomes can synthesize two types of melanin granules (eumelanin and brown melanin), and the ratio of the two melanin types determines animal skin and hair color (3). Mature melanin granules can protect skin keratinocytes from UV radiation damage (4). Melanosomes contain filamentous proteins and immunoactive substances, which are considered to be special lysosomes with antigen processing and presentation capabilities (2, 5). Thus, melanocytes not only are considered adenosine cells but also participate in the skin's immune system. Although many studies have been conducted on the mechanisms underlying melanogenesis, various aspects of the process remain unclear.

More than 100 genes have been identified to regulate melanin synthesis either directly or indirectly *via* the protein kinase C (6), Wnt (7, 8), and cAMP signaling pathways (9). α-Melanocyte-stimulating hormone (α-MSH) is an endogenous neuropeptide that binds with melanocortin 1 receptor (MC1R) on melanocytes to increase cyclic adenylate (cAMP) levels (10, 11). High levels of cAMP activate tyrosinase, which catalyzes tyrosine uptake from the blood by melanocytes and melanin granule synthesis *via* melanosomes (8, 12, 13). However, mechanisms underlying melanin synthesis and its roles in immunity remain unclear.

Long noncoding RNAs (lncRNA) are transcribed by RNA polymerase II and are multiexonic, polyadenylated, >200 nt in length, and located in the nucleus or the cytoplasm (14). They can be divided into antisense (antisense long noncoding RNAs), intronic transcript (intron noncoding RNAs), intergenic (lincRNAs), and promoter-associated or UTR-associated lncRNAs (15). lncRNAs in the cytoplasm can influence protein expression (16) and act as *cis*- and *trans*-acting elements to regulate gene expression (17, 18). Recent studies implicated lncRNAs in melanogenesis (19). The ceRNA network of lncRNA and circRNA has complex interactions involving ncRNA and mRNA related to skin and melanocyte development in mice. However, the role of these lncRNAs is poorly understood.

Although it has been confirmed that lncRNAs are widely involved in disease occurrence, cell metabolism, growth, development, and other physiological processes, their role in skin pigmentation in livestock animals has not been reported. In the present study, RNA transcription sequencing technology was used to explore the expression characteristics of lncRNAs and mRNAs in melanocytes isolated from goat skins with white and brown hairs, and the lncRNA–mRNA networks were constructed using functional prediction.

Materials and methods

Ethics approval

The study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of Anhui Agricultural University. The Institutional Animal Care and Use Committee of Anhui Agricultural University approved all animal surgeries (Approval no. 2016017).

Primary cell collection and isolation

The experimental animals were obtained from Linquan Comprehensive Experimental Station of Anhui Agricultural University. We used three female Boer goats that were 3 months old, and adequate drinking water and feed were provided throughout the experiment to keep them in healthy condition. Skin tissues (0.8 × 0.8 cm) with white and brown hair were isolated from the backs of the goats. The skin tissues were immersed in DMEM basal medium containing penicillin (400 U/ml) and streptomycin (400 μg/ml) and cut into thin strips (0.2 × 0.8 cm) before being placed in 0.25% Dispase II for over-digestion for 12–16 h after PBS cleaning. The epidermis and dermis were separated using ophthalmic tweezers, and the separated epidermis was cut into pieces and incubated in 0.25% trypsin+0.02% EDTA at 37°C for 10 min. An organic membrane filter (Solarbio, Beijing, China) was used to filter the digested products and collect the suspended cells. The cells were resuspended in 90% DMEM+10% FBS after centrifugation (1,000 rpm/min) and incubated at 37°C in a 5% CO_2 constant temperature incubator. After adherent growth, the cells were cultured in melanocyte medium (MelM, ScienCell Research Laboratories, Carlsbad, CA, USA) and then sub-cultured to the third generation for identification.

Identification of melanocytes

The morphological characteristics of melanocytes were microscopically observed (Leica, Buffalo Grove, USA). Dopamine staining (1 ml KH_2PO_4 + 3 ml Na_2HPO_4 + 12.5 ml 0.2% DoPa) and silver ammonia staining ($AgNO_3$ + $NH_3{\cdot}H_2O$) were used to detect the presence of melanin particles in the cytoplasm of melanocytes. Briefly, melanocytes were fixed in paraformaldehyde (4%) and washed with 1 × PBS (PH 7.4). The cells were immersed in dopamine and silver

ammonia stains and then incubated at 37°C for 1 h. Acidic calcium-binding protein (S100) was isolated and identified from bovine brains and is widely distributed in the cytoplasm of melanocytes, which are derived from neural crest cells (20). Immunohistochemical staining was used to detect the expression of S100 and other melanocyte markers in the cytoplasm. Melanocytes were fixed in paraformaldehyde (4%), incubated at 37°C in hydrogen peroxide (3%) for 15 min to block the action of endogenous peroxidases, and washed with 1 × PBS. The cells were immersed in bovine serum albumin (Rebiosci, Keilor, Australia) at 37°C for 25 min and incubated at 4°C overnight with rabbit primary antibodies against S100 (1:600, BioVision, San Francisco, USA). After the overnight incubation, the cells were washed with PBS (PH 7.4) and incubated with horseradish peroxidase-conjugated anti-rabbit IgG for 40 min at 37°C.

Cell proliferation assay

In this study, a CCK8 kit (Genomeditech, Shanghai, China) was used to detect the melanocyte proliferation rate. In triplicate, 3,000 melanocytes were seeded into 96-well plates with 100 μl of complete culture medium. The melanocytes were cultured for 0, 12, 24, 36, and 48 h before 10 μl of CCK8 buffer was added to 100 μl of medium and cultured again for 3 h. The culture plates were shaken for 10 min, and the optical density (OD) values were read at 450 nm.

Tyrosinase activity determination

Cells isolated from skins with white and brown hairs were added to 90 μl of 1% Triton X-100 and shaken for 5 min. Thereafter, 5 μl of 0.2% L-dopa was added, and the cells were incubated at 37°C for 30 min. The OD was measured at 490 nm using a microplate analyzer (Biotek, Vermont, USA), and tyrosinase activity (%) was calculated based on the OD of the treated and blank groups.

Melanin content measurement

Melanin content was analyzed using total alkali-soluble melanin (ASM) assay. Melanocytes were collected 72 h after transfection, rinsed three times with PBS (PH 7.4), and lysed in NaOH (1 mol/L) at 37°C for 45 min. Melanin content was measured using a Multiskan Spectrum microplate reader (Biotek, Vermont, USA) using an absorbance of 475 nm. The relative melanin content was normalized based on cell number.

Wound-healing assay

Melanocytes were seeded in 24-well cell culture plates coated with fibronectin (25 μg/well) until they grew to near confluence. A scratch was made using a pipette tip. Melanocytes were washed with PBS and cultured in an FBS-free medium. Cell migration was visualized under a microscope in a specific location at 0 h and 48 h after scratching.

RNA isolation and LncRNA-seq

RNA samples from melanocytes were extracted using TRIzol reagent. For sequence preparation, 5 μg of RNA from each sample was used as the input, and ribosomal RNA in the samples was removed using the rRNA Removal Kit (Aksomics, Beijing, China). cDNA libraries were generated using an RNA Library Kit (NEB, Beijing, China), and cDNA qualities were assessed using an Agilent Bioanalyzer 2100 system. Illumina Novaseq™ 6000 was used for sequencing after sample clustering. Using the specified GCF/001 genome/*Capra hircus* (goat)/2016 (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.2) as a reference (21), sequence alignment and subsequent analyses were performed. Pearson correlation coefficients (22) were calculated to reflect the degree of linear correlation between two groups of samples. The StringTie transcript assembly software (23) was used to assemble the reads. The known mRNA and transcripts < 200 bp were removed, and lncRNA prediction was performed for the remaining transcripts. The coding potential calculator (CPC) (24) and the Coding-Non-Coding Index (CNCI) (25) were used as prediction software.

Expression pattern analysis of mRNA and lncRNA

The position of the lncRNA in the genome was determined, and classified statistics for lncRNA were generated according to the comparative analysis between lncRNA sequences and the reference genome. Differential expression of mRNAs and lncRNAs between melanocytes from skins with white and brown hairs was analyzed using false discovery rate (FDR) and fold change (26). An FPKM was developed to measure the expression of genes and lncRNAs to overcome the isoform length-dependence of the read counts (27).

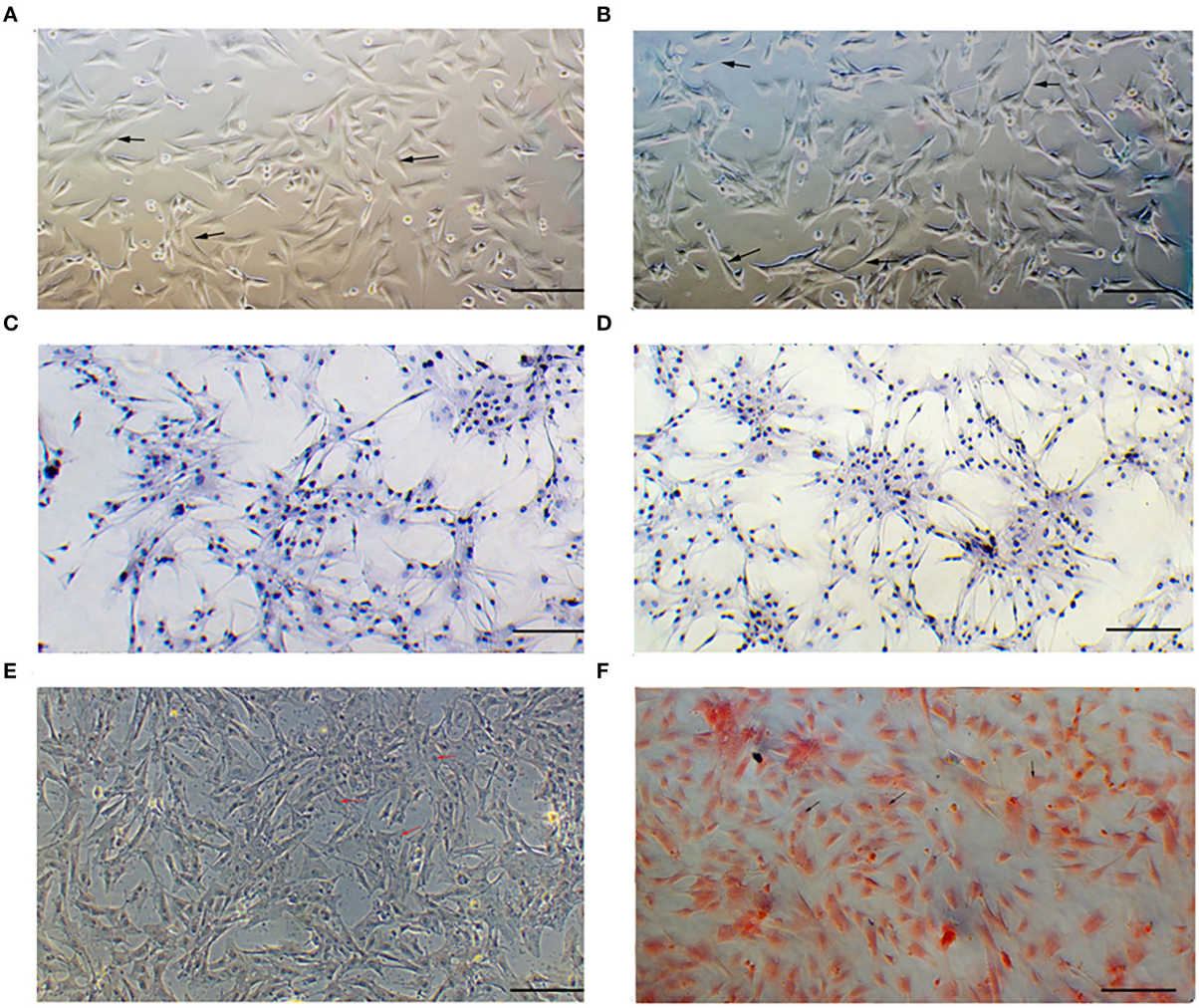


FIGURE 1
Identification of melanocytes cultured *in vitro*. **(A)** Morphology of cells isolated from skins with white hair as observed under a microscope. **(B)** Morphology of cells isolated from skins with brown hair as observed under a microscope. **(C,D)** S100 protein signaling detected by cellular immunohistochemical staining. **(E)** Melanin granules observed using dopamine staining. **(F)** Melanin granules observed using silver ammonia staining. Bar = 0.1 mm.

Functional analysis of differentially expressed mRNAs (DEGs) and lncRNAs (DELs)

As post-transcriptional products, lncRNAs can form miRNAs *via* shearing (28). Therefore, homology analysis was used to compare whether the DELs have miRNA precursor structures. lncRNAs play crucial roles as enhancers (29) and promoters (30) to regulate adjacent coding genes. lncRNAs can regulate the expression of genes that overlap with the lncRNA transcription start or end site at distances < 100 kb (31). In the present study, we analyzed the location of DELs and mRNAs and predicted their relationship. The Gene Ontology (GO) and KEGG databases were used to analyze the functions of the DEGs and the target genes of DELs.

Statistical analysis

The differences in cell proliferation, tyrosinase activity, cell migration, and melanin amount were determined using Fisher's protected LSD test and ANOVA. The research data were analyzed using SPSS 11.5 (Chicago, IL, USA). The results were expressed as mean ± standard deviation (SD), with a P-value < 0.05 indicating statistical significance.

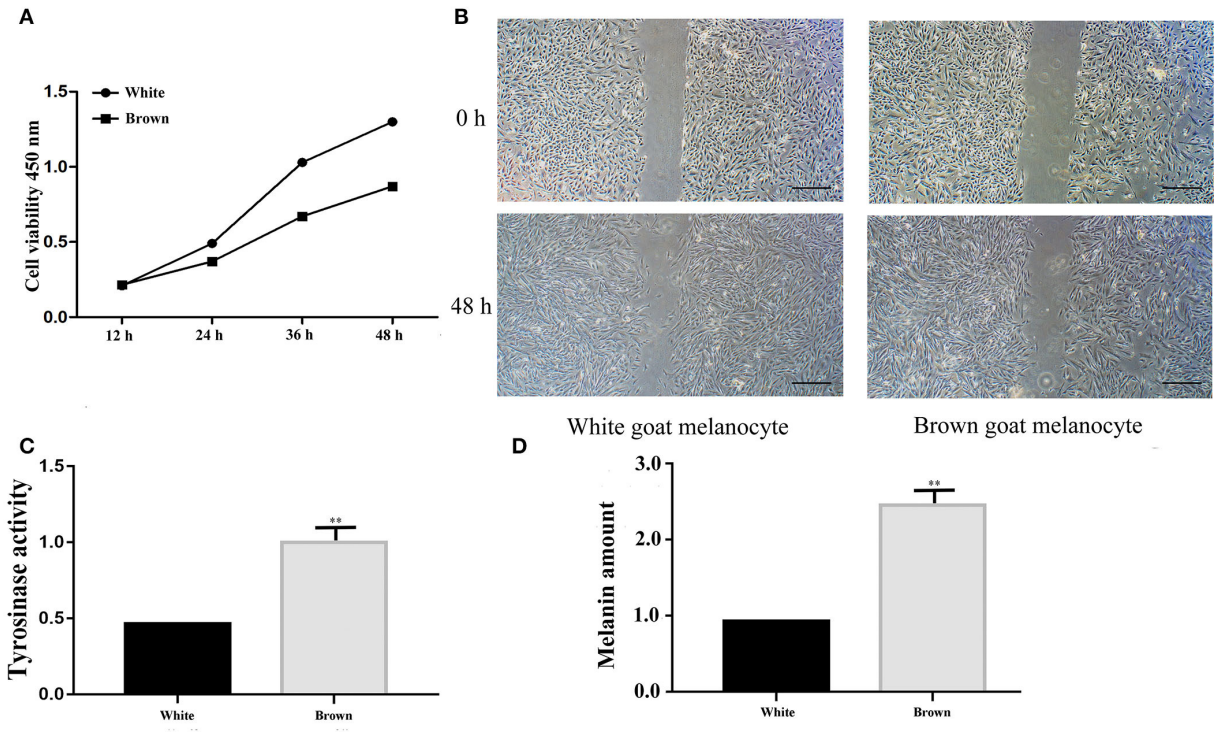


FIGURE 2
Identification of melanocyte phenotypes isolated from Boer goat skins with white and brown hairs. **(A)** Proliferation rates of melanocytes as detected by a CCK8 assay. **(B)** Migration rates of melanocytes as detected by the wound-healing assay. **(C)** Tyrosinase activity of melanocytes as detected by L-dopa staining. **(D)** Melanin production of melanocytes as detected by an ASM assay. Data are shown as the means ± SD of relative fold-change ($n = 3$ per group). $^{**}P < 0.01$. Bar = 0.25 mm.

Results

Identification of melanocytes

In the microscopic field, the cells isolated from Boer goat skins with white (Figure 1A) or brown (Figure 1B) hairs were found to be fusiform-shaped, similar to the general morphological structure of melanocytes. Immunohistochemical staining showed that S100 is detected in the cytoplasm of the cells (Figures 1C,D). Dopamine (Figure 1E) and silver ammonia staining (Figure 1F) showed melanin granules in the cytoplasm of cells.

Melanocyte phenotypes

The cell proliferation assay showed that the proliferation rates of melanocytes isolated from Boer goat skins with white hair were higher than those with brown hair (Figure 2A). The migration rates of melanocytes in skins with white hair were also higher (Figure 2B). Tyrosinase activity ($^{**}P < 0.01$; Figure 2C), as well as melanin production rates ($^{**}P < 0.01$; Figure 2D), was significantly higher in melanocytes isolated from skins with brown hair.

Expression patterns of lncRNAs and mRNAs

When comparing melanocytes isolated from Boer goat skins with white and brown hairs, Pearson's correlation coefficient was 0.969, indicating a marked correlation (Figure 3A). Quality control analysis found that the expression density map of each sample followed a normal distribution, and the expression trends of biological replicates tended to be consistent (Figures 3B,C). Transcriptional analysis showed that there were 27,361 lncRNAs and 54,128 mRNAs in the two cell groups, and in melanocytes, the expression levels of mRNAs were significantly higher than those of lncRNAs (Figure 3D). When compared with mRNAs, most lncRNAs only had one exon, which may be due to the process of lncRNA formation (Figure 3E). Statistics showed that most lncRNAs in melanocytes were derived from exons and introns, with a small portion being derived from intergene regions (Figure 3F).

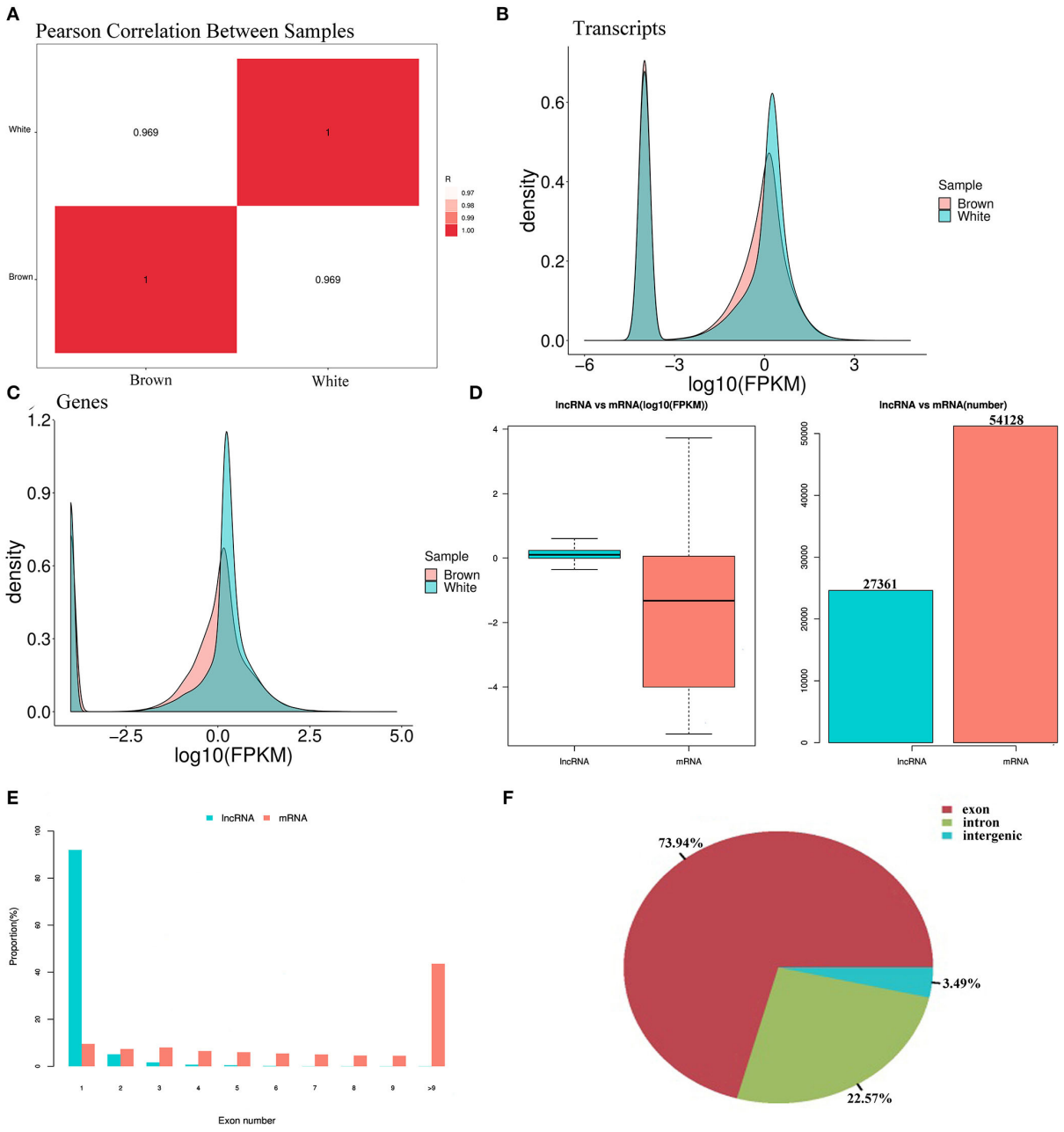


FIGURE 3
Expression patterns of lncRNA and mRNA in melanocytes isolated from Boer goat skins with white and brown hairs. **(A)** Pearson's correlation coefficient between different samples. **(B,C)** Expression value density of transcripts and gene expression trend. **(D)** Expression levels of lncRNAs and mRNAs in melanocytes. **(E)** Distribution of lncRNA and mRNA length in melanocytes. **(F)** Distribution of lncRNA located in melanocytes.

Divergent expression patterns of mRNA

Transcriptome sequencing analysis found 2,076 DEGs between melanocytes isolated from goat skins with white and brown hairs (Figures 4A,B); the top 30 DEGs are shown in Supplementary Table S1. The KEGG database was used to annotate the 2,076 DEGs, and the top 20 KEGG annotation terms (KEGG term) were selected to generate a scatter diagram of the KEGG enrichment analysis (Figure 4C). The analysis found that the DEGs were significantly enriched in the ErbB signaling (32) and focal adhesion regulatory pathways (33), which

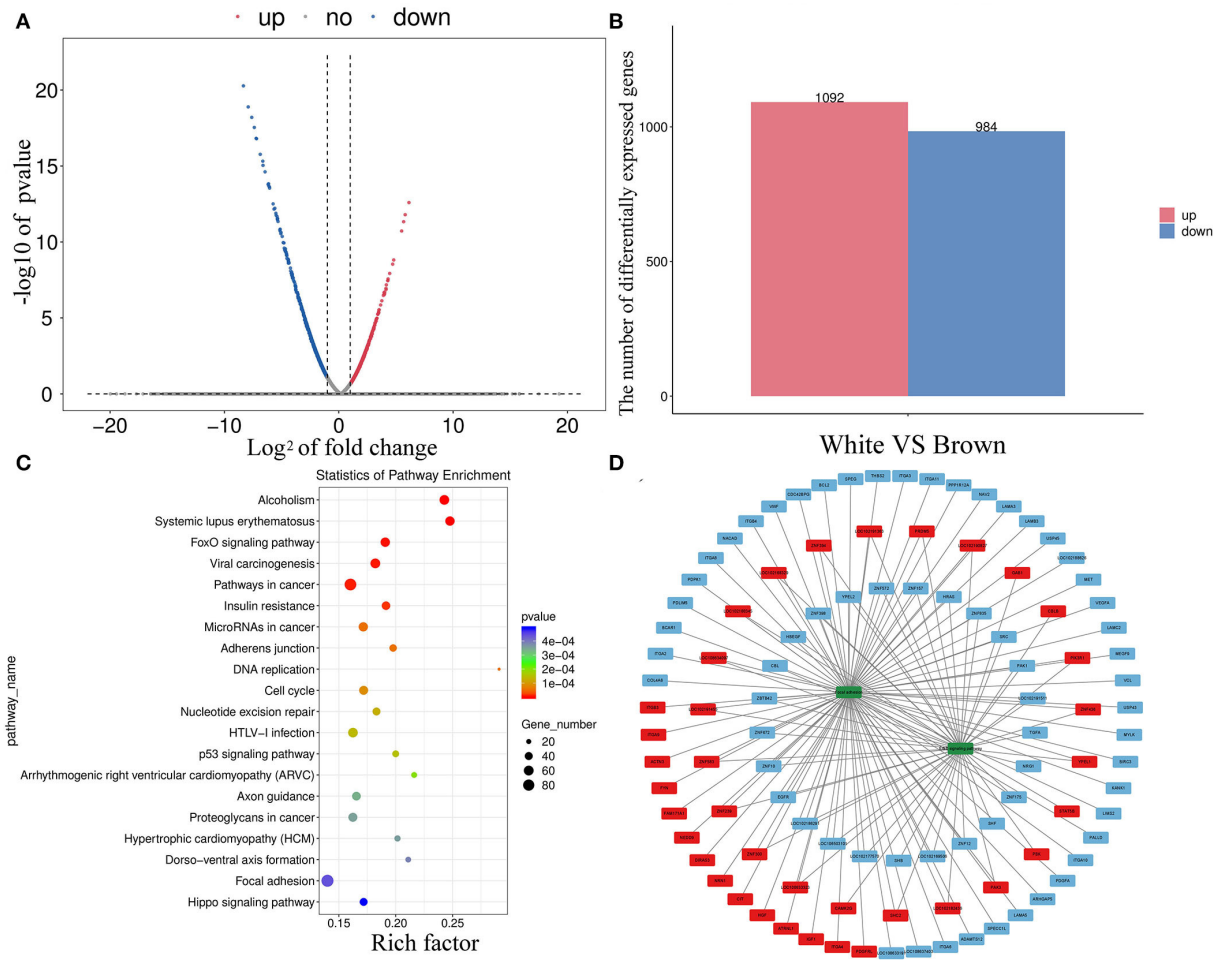


FIGURE 4
Divergent expression patterns of mRNAs in melanocytes isolated from Boer goat skins with white and brown hairs. **(A,B)** Divergent expression patterns of mRNAs in melanocytes. **(C,D)** DEG annotation results were obtained using the KEGG database; the red symbol marks positive regulatory genes, the blue marks negative regulatory genes, and the green marks key signaling pathway.

are involved in tyrosinase activation and cell migration (Figure 4D).

Divergent expression patterns of lncRNAs

In total, 1,536 lncRNAs (251 were downregulated and 1,285 were upregulated) were differentially expressed in melanocytes isolated from goat skins with white and brown hairs (Figures 5A,B; Supplementary Table S2). Homology analysis showed that there were nine DELs with the same amino acid sequences as miRNA precursors and may correspond to miRNA precursors (Supplementary Table S3). lncRNA also has a potential *cis*-regulatory capacity for adjacent coding genes (34, 35). After the prediction of DEL *cis*-target genes (Supplementary Table S4), statistics from KEGG annotation found that the top potential target gene-enriched terms were involved in GABAergic synapse (Figure 5C), which may promote the expression and secretion of gonadotropin signaling molecules (36). Moreover, KEGG annotation revealed that the potential target genes of DELs were involved in the cAMP (37), MAPK (38), and ErbB signaling pathways (32), as well as other pigmentation-related pathways (Figure 5D).

Identification of the core lncRNA–mRNA networks

Core networks were identified from common terms of *cis*-regulation of DELs and DEMs to analyze the molecular mechanisms underlying cell proliferation, migration, and melanogenesis of melanocytes. Statistics found that 136 *cis*-genes of DELs overlapped with DEMs (Figure 6A; Supplementary Table S5), and 139 DELs and *cis*-genes

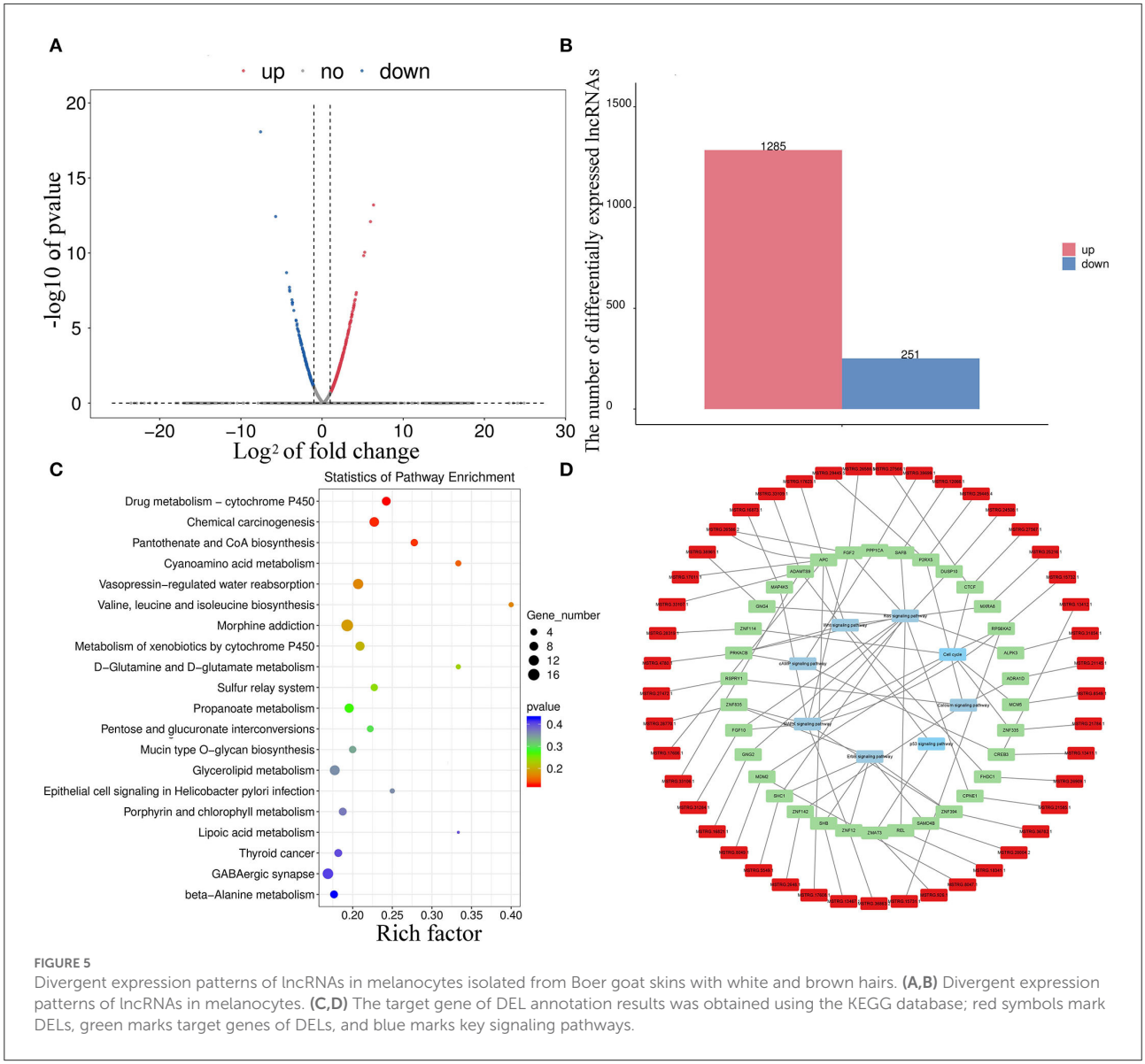


FIGURE 5
Divergent expression patterns of lncRNAs in melanocytes isolated from Boer goat skins with white and brown hairs. **(A,B)** Divergent expression patterns of lncRNAs in melanocytes. **(C,D)** The target gene of DEL annotation results was obtained using the KEGG database; red symbols mark DELs, green marks target genes of DELs, and blue marks key signaling pathways.

overlapped with DEMs with the same expression trends. KEGG annotation found that the top enriched terms of target DEMs were involved in the MAPK signaling pathway (Figure 6B), and thus, the core DEL–DEM networks were constructed (Figures 6C,D).

Discussion

In this study, melanocytes from goat skins were successfully isolated and verified *in vitro*. The morphological analysis found that melanocytes isolated from goat skins with white hair were spindle-shaped with a short protrusion, while the number and the length of protrusions in melanocytes isolated from skins with brown hair were increased, which may be related to melanin transport.

Phenotypic testing revealed that melanocytes isolated from goat skins with white and brown hairs showed significant differences in proliferation, migration, and melanogenesis. To analyze the molecular mechanisms underlying phenotypic differences in the two groups, we determined the expression profiles of lncRNAs and mRNAs. According to the transcriptome data analysis, there were 2,076 DEMs in melanocytes isolated from goat skins with white and brown hairs. This suggests that these DEGs were possibly related to regulating melanocyte proliferation, migration, and melanogenesis. The ErbB signaling pathway (32) and the focal adhesion regulatory pathway (33) are key for regulating tyrosinase activation and cell migration,

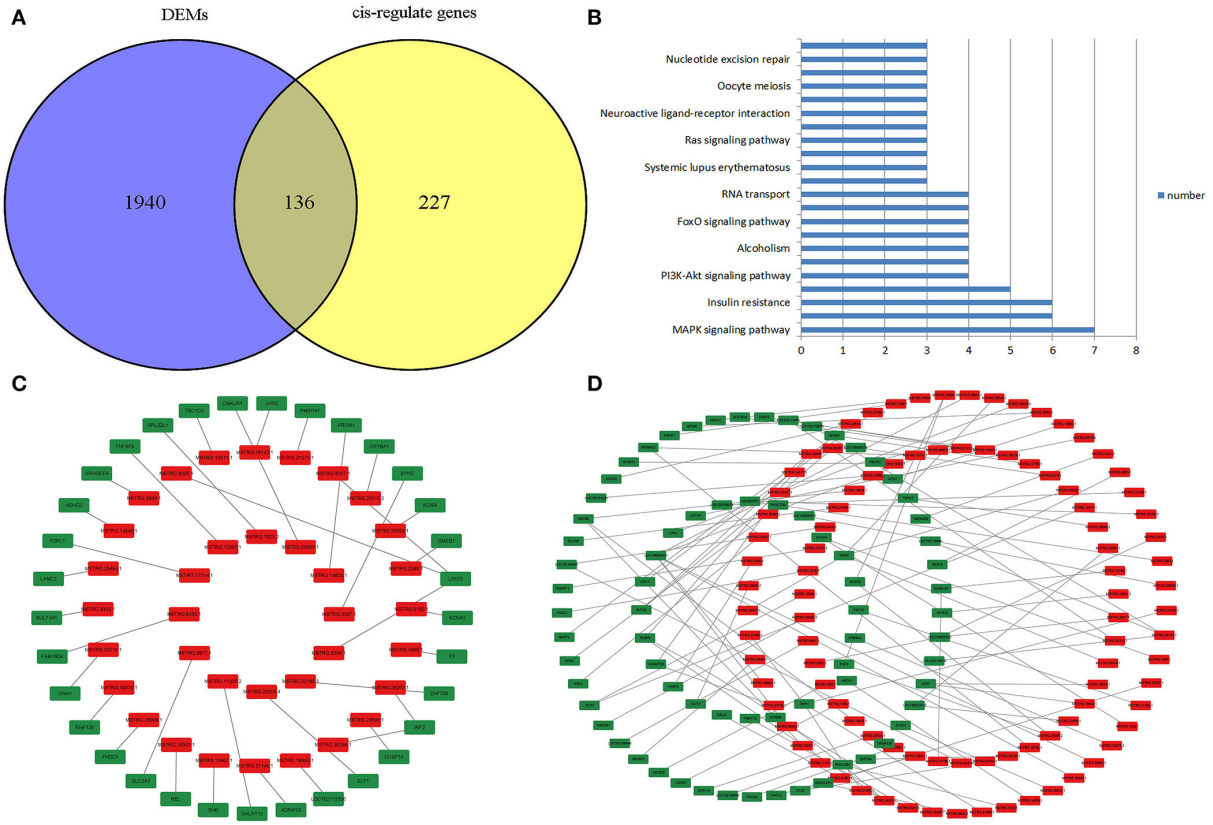


FIGURE 6
Identification of the core lncRNA–mRNA network. **(A)** Venn plot of *cis*-genes of DELs overlapped with DEMs. **(B)** Functional enrichment analysis of DELs that target DEMs. **(C)** Core negative regulatory DEL–DEM networks in melanocytes. **(D)** Core positive regulatory DEL–DEM networks in melanocytes; the red symbol marks the DELs, and green marks the DEMs.

and the KEGG annotation results showed that there were multiple DEGs located in these two regulatory pathways. For example, ITGA6 is significantly overexpressed in hepatocellular carcinoma and mediates tumor progression (39). ESRRG may promote the proliferation and migration of Ishikawa cells (40). FGF2 stimulates the growth of trunk neural crest cells and promotes melanocytic commitment (41). Thus, DEGs can provide novel insights for analyzing the characteristics of melanocytes.

Long noncoding RNAs are associated with multiple physiological processes in animals [(21); 18]. In the present study, we found 27,361 novel potential lncRNAs in melanocytes. The expression of lncRNAs usually causes tissue- and stage-specific patterns in animals (30). Thus, the abundance profiles identified were assumed to be candidate melanocyte-specific lncRNAs. Currently, the functions of lncRNAs are investigated through their target genes, which may be regulated by *cis*-regulatory methods as previously described (34). In the present study, we predicted the *cis*-target genes of DELs and annotated the functions of potential target genes using the KEGG databases. The annotation results found that multiple predicted target genes were involved in MAPK signaling and the regulation of the actin cytoskeleton signaling pathway, which are closely related to cell proliferation, migration, and Melanogenesis. For example, FGF10 may upregulate cell proliferation in white adipose tissues (42), and APCs are widely involved in physiological processes such as cell proliferation, migration, and melanin production (43). These results suggest that the related DELs may participate in the physiological processes by regulating target genes. lncRNAs may also act as miRNA precursors to develop mature miRNAs and exert their function through targeted adsorption of downstream genes (44). The comparison of DEL sequences revealed that nine DELs had high homology with miRNA precursors, which may be their respective precursors. It has been reported that let-7e-5p and miR-574-5p are involved in the proliferation and migration of a variety of cells (45).

The *cis*-regulatory effects of lncRNAs on neighboring genes have been well-examined (30, 35). In this study, 136 *cis*-genes of DELs overlapped with DEMs, and 139 paired DELs and *cis*-genes had the same expression

trends. KEGG annotation statistics found that the top enriched terms of the target DEMs were involved in the MAPK signaling pathway, which is the key pathway for regulating proliferation, migration, and melanogenesis, and thus, the DEL-DEM networks were constructed.

Conclusion

In the study, melanocytes were successfully isolated from Boer goat skins *in vitro*. Phenotypic testing revealed that melanocytes isolated from goat skins with white and brown hair showed significant differences in cell proliferation, migration, and melanogenesis ($^{**}P < 0.01$). Through prediction analysis, genome-wide stage-specific candidate lncRNAs were identified in goat melanocytes. DELs and DEGs in melanocytes isolated from goat skins were screened, and the functions of the source genes of the DELs and DEMs were annotated. The functions of DELs in *cis*-regulation and miRNA precursor functions were investigated. Based on our results, multiple lncRNA–mRNA networks may be involved in key signaling pathways in melanocyte proliferation, migration, and melanin production. This research provided novel bioinformatic insights into the roles of lncRNAs in mammalian pigmentation.

Data availability statement

The datasets for this study can be found in the NCBI database and the BioProject accession number is PRJNA865095.

Ethics statement

The animal study was reviewed and approved by the Ethics Committee of Anhui Agricultural University.

Author contributions

JK-y and ZY-h: conceptualization, investigation, data analysis, and writing-original draft preparation. WR-j, IK, and ZY-W: methodology. All authors have read and approved the final manuscript.

Funding

This study was funded by the Anhui Provincial Natural Science Foundation (grant number 2008085QC158), the University Natural Science Research Project of Anhui Province (grant number KJ2019A0165), and Anhui Provincial Natural Science Foundation (grant number 1908085QC144).

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/fvets.2022.1009174/full#supplementary-material

References

1. Nishimura EK, Jordan SA, Oshima H, Yoshida H, Osawa M, Moriyama M, et al. Dominantrole of the niche in melanocyte stem-cell fate determination. *Nature*. (2002) 416:854–60. doi: 10.1038/416854a

2. Koike S, Yamasaki K, Yamauchi T, Shimada-Omori R, Tsuchiyama K, Ando H, et al. TLR3 stimulation induces melanosome endo/phagocytosis through RHOA and CDC42 in human epidermal keratinocyte. *J Dermatol Sci*. (2019) 96:168–77. doi: 10.1016/j.jdermsci.2019.11.005

3. Hoffmann G, Schobersberger W, Rieder J, Smolny M, Sepp N. Human dermal microvascular endothelial cells express inducible nitric oxide synthase *in vitro*. *J Invest Dermatol*. (1999) 12:387–90. doi: 10.1046/j.1523-1747.1999.00505.x

4. Supp DM, Hahn JM, Lloyd CM, Combs KA, Swope VB, Zalfa AM, et al. Light or dark pigmentation of engineered skin substitutes containing melanocytes protects against UV-induced DNA damage *in vivo*. *J Burn Care Res*. (2020) 41:S78–9. doi: 10.1093/jbcr/iraa029

5. Ahn JH, Park TJ, Jin SH, Kang HY. Human melanocytes express functional toll-like receptor 4. *Exp Dermatol*. (2008) 17:412–7. doi: 10.1111/j.1600-0625.2008.00701.x

6. Jothy SL, Saito T, Kanwar JR, Chen Y, Aziz A, Leong YH, et al. Radioprotective activity of Polyalthia longifolia standardized extract against X-ray radiation injury in mice. *Phys Med*. (2016) 32:150–61. doi: 10.1016/j.ejmp.2015.10.090

7. Yang SS, Liu B, Ji KY, Fan RW, Dong CS. MicroRNA-5110 regulates pigmentation by co-targeting melanophilin and WNT family member 1. *FASEB J*. (2018) 32:5405–12. doi: 10.1096/fj.201800040R

8. Zou DP, Chen YM, Zhang LZ, Yuan XH, Chen J. SFRP5 inhibits melanin synthesis of melanocytes in vitiligo by suppressing the Wnt/β-catenin signaling. *Genes Dis.* (2020) 8:677–88. doi: 10.1016/j.gendis.2020.06.003

9. Holcomb NC, Bautista RM, Jarrett SG, Carter KM, Gober MK, D'Orazio JA. cAMP-mediated regulation of melanocyte genomic instability: a melanoma-preventive strategy. *Adv Protein Chem Struct Biol.* (2018) 115:247–95. doi: 10.1016/bs.apcsb.2018.10.008

10. Jose CGB, Zalfa AM, Celia JC. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. *Pigm Cell Melanoma R.* (2014) 27:699–720. doi: 10.1111/pcmr.12257

11. Jackson E, Heidl M, Imfeld D, Meeus L, Schuetz R, Campiche R. Discovery of a highly selective MC1R agonists pentapeptide to be used as a skin pigmentation enhancer and with potential anti-aging properties. *Int J Mol Sci.* (2019) 20:6143. doi: 10.3390/ijms20246143

12. Seung EL, See-Hyoung P, Sae WO, Yoo JA, Kwon K, Park SJ, et al. Beauvericin inhibits melanogenesis by regulating cAMP/PKA/CREB and LXR-α/p38 MAPK-mediated pathways. *Sci Rep.* (2018) 8:14958. doi: 10.1038/s41598-018-33352-8

13. Cheli Y, Ohanna M, Ballotti R, Bertolotto C. Fifteen-year quest for microphthalmia-associated transcription factor target genes. *Pigm Cell Melanoma R.* (2010) 23:27–40. doi: 10.1111/j.1755-148X.2009.00653.x

14. Kertesz M, Yue W, Mazor E, Rinn JL, Nutter RC, Chang HY, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature.* (2010) 467:103–7. doi: 10.1038/nature09322

15. Ulitsky I, Bartel D. lincRNAs: genomics, evolution, and mechanisms. *Cell 154.* (2013) 26–46. doi: 10.1016/j.cell.2013.06.020

16. Zhou J, Fan Y, Chen H. Analyses of long non-coding RNA and mRNA profiles in the spinal cord of rats using RNA sequencing during the progression of neuropathic pain in an SNI model. *RNA Biol.* (2017) 14:1810–26. doi: 10.1080/15476286.2017.1371400

17. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature.* (2016) 539:452–5. doi: 10.1038/nature20149

18. National Center for Biotechnology Information. *Capra Hircus.* (2016). Available online at: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_ 001704415.2_ARS1/ (accessed August 29, 2016).

19. Zhu Z, Ma Y, Li Y, Li P, Tang Z. The comprehensive detection of miRNA, lncRNA, and circRNA in regulation of mouse melanocyte and skin development. *Biol Res.* (2020) 53:4. doi: 10.1186/s40659-020-0272-1

20. Shirley SH, Maltzan VK, Robbins PO, Kusewitt DF. Melanocyte and melanoma cell activation by calprotectin. *J Skin Cancer.* (2014) 2014:846249. doi: 10.1155/2014/846249

21. Irina AE, Vladimir SS. The Role of cis- and trans-acting RNA regulatory elements in leukemia. *Cancers.* (2020) 12:3854. doi: 10.3390/cancers12123854

22. Gerhard, N. *Dictionary of Pharmaceutical Medicine*. Cham: European Center of Pharmaceutical Medicine (2009). p. 132.

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

24. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* (2016) 11:1650–67. doi: 10.1038/nprot.2016.095

25. 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

26. Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage.* (2009) 44:62–70. doi: 10.1016/j.neuroimage.2008.05.021

27. Mortazavi A, Williams BA, Mccue K, Schaeffer L. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* (2008) 5:621–8. doi: 10.1038/nmeth.1226

28. Wu Q, Guo L, Jiang F, Li L, Li Z, Chen F. Analysis of the miRNA–mRNA–lncRNA networks in ER+ and ER breast cancer cell lines. *J Cell Mol Med.* (2015) 19:2874–87. doi: 10.1111/jcmm.12681

29. Soibam B. Super-lncRNAs: identification of lncRNAs that target super-enhancers via RNA:DNA:DNA triplex formation. *RNA.* (2017) 23:1729–42. doi: 10.1261/rna.061317.117

30. Yan P, Luo S, Lu JY, Shen X. Cis-and trans-acting lncRNAs in pluripotency and reprogramming. *Curr Opin Genet Dev.* (2017) 46:170–8. doi: 10.1016/j.gde.2017.07.009

31. Wang L, Zhao Y, Bao X, Zhu X, Kwok KY, Sun K, et al. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res.* (2015) 25:335–50. doi: 10.1038/cr.2015.21

32. Zhang K, Wong P, Salvaggio C, Salhi A, Osman I, Bedogni B. Synchronized targeting of notch and ERBB signaling suppresses melanoma tumor growth through inhibition of Notch1 and ERBB3. *J Invest Dermatol.* (2016) 136:464–72. doi: 10.1016/j.jid.2015.11.006

33. Schringer K, Maxeiner S, Schalla C, Rütten S, Zenke M, Sechi A. LSP1-myosin1e bimolecular complex regulates focal adhesion dynamics and cell migration. *FASEB J.* (2021) 35:e21268. doi: 10.1096/fj.202000740RR

34. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell.* (2005) 123:1279–91. doi: 10.1016/j.cell.2005.11.035

35. Stewart GL, Sage AP, Enfield KSS, Marshall EA, Wan LL. Deregulation of a Cis-acting lncRNA in non-small cell lung cancer may control HMGA1 expression. *Front Genet.* (2021) 11:615378. doi: 10.3389/fgene.2020.615378

36. Farkas I, Kalló I, Deli L, Vida B, Hrabovszky E, Fekete C, et al. Retrograde endocannabinoid signaling reduces GABAergic synaptic transmission to gonadotropin-releasing hormone neurons. *Endocrinology.* (2010) 151:5818–29. doi: 10.1210/en.2010-0638

37. Bang J, Zippin JH. Cyclic adenosine monophosphate (cAMP) signaling in melanocyte pigmentation and melanomagenesis. *Pigm Cell Melanoma R.* (2020) 34:28–43. doi: 10.1111/pcmr.12920

38. Ohta H, Yabuta Y, Kurimoto K, Nakamura T, Murase Y, Yamamoto T, et al. Cyclosporin A and FGF signaling support the proliferation/survival of mouse primordial germ cell-like cells in vitro. *Biol Reprod.* (2021) 104:344–60. doi: 10.1093/biolre/ioaa195

39. Yang X, Song D, Zhang J, Feng H, Guo J. PRR34-AS1 sponges miR-498 to facilitate TOMM20 and ITGA6 mediated tumor progression in HCC. *Exp Mol Pathol.* (2021) 120:104620. doi: 10.1016/j.yexmp.2021.104620

40. Yuan B, Liang Y, Wang D, Luo F. MiR-940 inhibits hepatocellular carcinoma growth and correlates with prognosis of hepatocellular carcinoma patients. *Cancer Sci.* (2015) 106:819–24. doi: 10.1111/cas.12688

41. Teixeira B, Amarante-Silva D, Visoni SB, Garcez RC, Trentin A. FGF2 stimulates the growth and improves the melanocytic commitment of trunk neural crest cells. *Cell Mol Neurobiol.* (2020) 40:383–93. doi: 10.1007/s10571-019-00738-9

42. Konishi M, Asaki T, Koike N, Miwa H, Itoh N. Role of Fgf10 in cell proliferation in white adipose tissue. *Mol Cell Endocrinol.* (2006) 249:71–7. doi: 10.1016/j.mce.2006.01.010

43. Nakagome K, Dohi M, Okunishi K, Komagata Y, Nagatani K, Tanaka R, et al. In vivo IL-10 gene delivery suppresses airway eosinophilia and hyperreactivity by down-regulating APC functions and migration without impairing the antigen-specific systemic immune response in a mouse model of allergic airway inflammation. *J Immunol.* (2005) 174:6955–66. doi: 10.4049/jimmunol.174.11.6955

44. Wang PF, Dai LM, Ai J, Wang YM, Ren FS. Identification and functional prediction of cold-related long non-coding RNA (lncRNA) in grapevine. *Sci Rep.* (2019) 9:6638. doi: 10.1038/s41598-019-43269-5

45. Wang S, Jin S, Liu MD, Pang P, Wu H, Qi ZZ, et al. Hsa-let-7e-5p inhibits the proliferation and metastasis of head and neck squamous cell carcinoma cells by targeting chemokine receptor 7. *J Cancer.* (2019) 10:1941–8. doi: 10.7150/jca.29536

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fvets.2022.1078394

Check for updates

OPEN ACCESS

EDITED BY
Hongyu Liu,
Anhui Agricultural University, China

REVIEWED BY
Yong Liu,
Fuyang Normal University, China
Karima Mahmoud,
National Research Centre, Egypt
Yaokun Li,
South China Agricultural University, China

*CORRESPONDENCE
Eryao Wang
✉ eryaowang@outlook.com
Shijie Lyu
✉ sjlyu@outlook.com

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

RECEIVED 24 October 2022
ACCEPTED 05 December 2022
PUBLISHED 20 December 2022

CITATION
Zhai Y, Shi Q, Chu Q, Chen F, Feng Y, Zhang Z, Qi X, Arends D, Brockmann GA, Wang E and Lyu S (2022) miRNA profiling in intrauterine exosomes of pregnant cattle on day 7. *Front. Vet. Sci.* 9:1078394. doi: 10.3389/fvets.2022.1078394

COPYRIGHT
© 2022 Zhai, Shi, Chu, Chen, Feng, Zhang, Qi, Arends, Brockmann, Wang and Lyu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

miRNA profiling in intrauterine exosomes of pregnant cattle on day 7

Yaying Zhai[1,2], Qiaoting Shi[1], Qiuxia Chu[1], Fuying Chen[1], Yajie Feng[1], Zijing Zhang[1], Xinglei Qi[3], Danny Arends[4], Gudrun A. Brockmann[5], Eryao Wang[1]* and Shijie Lyu[1,6]*

[1]Institute of Animal Husbandry and Veterinary Science, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China, [2]College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou, China, [3]Center of Animal Husbandry Technical Service in Biyang, Zhumadian, China, [4]Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom, [5]Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Humboldt-Universität zu Berlin, Berlin, Germany, [6]The Shennong Laboratory, Zhengzhou, Henan, China

Intrauterine exosomes have been identified to be involved in the embryo development and implantation. The aim of this study was to explore the role of miRNAs in intrauterine exosomes in bovine pregnancy. Intrauterine exosomes were collected from uterine flushing fluids of three donor and three recipient Xianan cows 7 days after fertilization. Intrauterine exosomes miRNAs were extracted and the exosomal miRNAs expression levels were analyzed. Sixty miRNAs differed significantly in their amounts between donors and recipients (p-$value < 0.05$, $|\log2(\text{FoldChange})| > 1$). Twenty-two miRNAs were upregulated and 38 downregulated in the group of donor cows. The bta-miR-184 was the most significant ($P_{\text{Benjamini-Hochberg}} < 0.001$). A total of 9,775 target genes were predicted using the 60 miRNAs. GO and KEGG analysis showed that the target genes were enriched in several biological processes or pathways associated with embryo implantation and endometrial development, such as cell adhesion, cell junction, focal adhesion, and Rap1 signaling pathway. Our findings suggest that, in cattle early pregnancy stage, these differently expressed miRNAs in intrauterine exosomes involved in embryo implantation and endometrial development, which may exert a significant effect and influence the uterine microenvironment for embryo implantation. These results could provide reference for screening and exploring the intrauterine exosomal miRNA affecting embryo implantation.

KEYWORDS

bovine, uterine flushing, receptive endometrium, apoptosis, proliferation

Introduction

Low conception rate is a problem which significantly reduces the economic benefits of cattle industry. It is reported that more than 50% embryo loss occurs during the first week of pregnancy (1). The establishment of pregnancy requires not only the normal development of the embryo itself but also a favorable environment for embryo implantation (2). During the implantation stage of bovine embryos, uterine fluid supports embryonic development and survival (3). To enlighten potential causes for

bovine low conception rate, it is necessary to better understand the interaction between the early embryo and the uterine fluid.

Exosomes are 50–150 nm extracellular vesicles secreted by various cell types. They are a media for intercellular communication (4, 5). They contain nucleic acids, proteins, and lipids, and play an important role in adhesion, proliferation, apoptosis, inflammation, and immune response in recipient cells (6, 7). Growing evidence has shown that exosomes are important for the establishment of pregnancy *via* affecting oogenesis, oocyte maturation and fertilization, embryo-maternal cross talk, and embryo implantation (7, 8).

In uterine flushing fluids (UFs), exosomes have been identified and found to be involved in the embryo development and implantation as well (9, 10). In sheep, endometrial exosomes regulate the secretion of interferon tau (IFNT), a pregnancy recognition signal in ruminants (11). In cattle, treatment of endometrial epithelial cells with exosomes isolated from the UFs of pregnant cattle at day 17 increased the expression of apoptosis-related genes and IFNT-stimulated genes (10, 12). Exosomes are particularly enriched in miRNAs, which can regulate the translation of mRNA to protein in the recipient cells. Thereby, exosomes with their content are most likely factors influencing the establishment of pregnancy (13). For example, in humans, miR-30d from exosomes in the uterine lumen is increased during the window of implantation. Mouse *in vitro* experiments showed that exosome-associated miR-30d was taken up by the trophectoderm of blastocysts and which contributed to increase trophectoderm adhesion (14). In goats, exosomal miRNAs in UFs were found to be required for uterine receptivity and embryo implantation (15).

After fertilization, the bovine morula enters the uterus on days 4–5 where it forms a blastocyst by day 7. We hypothesize that miRNAs of exosomes in the intrauterine fluid between day 5 and 7 after fertilization might be required for the establishment of a favorable uterus environment for the embryo implantation on day 7. Therefore, we aim to explore potential functions of miRNAs in intrauterine exosomes for the implantation of bovine embryos. In this study, the diversity and amounts of miRNAs in the exosomes of intrauterine fluids of donor and recipient cows on day 7 after fertilization were investigated. The result of this study could provide reference for screening and exploring the intrauterine exosomal miRNA affecting embryo implantation, and provide the basis for further clarification of intrauterine fluids exosomes in the early pregnancy regulation of cattle.

2. Materials and methods

2.1. Collection of bovines flushed uterine fluids

All animal procedures in this study were performed in accordance with the guidelines of the Committee for Experimental Animals at Henan Academy of Agricultural Sciences. Ten Xianan cows (a beef cattle breed in China) were selected as donors for synchronous estrus and superovulation. The hormones used for superovulation are follicle-stimulating hormone (FSH) and prostaglandin (PG), which are administered by intramuscular injection. FSH is administered over 4 consecutive days, twice daily in decreasing doses. Over the 4 days, FSH is administered every 12 h, at 7 am and 7 pm. The treatment protocol is as follows: day 1, 70 mg every 12 h; day 2, 50 mg every 12 h; day 3, 30 mg every 12 h; day 4, 20 mg every 12 h. At 7 pm on day 3 of FSH treatment, 0.1 mg PG was injected and repeated 12 h later. Estrus can be expected 36–48 h later. The artificial insemination was performed after superovulation. Another 23 Xianan cows were selected as recipients for synchronous estrus. On the 7th day after artificial insemination of donors, 100 ml bovine UFs were collected by uterine flushing using sterile Gibco Minimum Essential Media with 0.7% 200 mM Tris buffer (pH = 7.2). Three donor cows with well-developed blastocysts were used for further analysis (Table 1). Three recipient cows having similar body weight and age were selected as matched controls. All UF samples were stored at −80°C until further analysis.

TABLE 1 The information of the donor and recipient cows.

Individual	Age (year)	Body weight (kg)	Number of blastocysts obtained
Donor cow No.1	4	602	4
Donor cow No.2	3	486	2
Donor cow No.3	6	658	4
Recipient cow No. 1	4	530	–
Recipient cow No. 2	4	482	–
Recipient cow No. 3	6	610	–

2.2. Isolation of exosomes from uterine flushing fluids

Each UF sample was thawn at 37°C and centrifuged at 2,000 g for 30 min at 4°C. The supernatant was collected and then centrifuged at 12,000 g for 45 min at 4°C. After filtering through a 0.45 μm filter (Millipore, USA), the resulting supernatant was centrifuged at 110,000 g for 70 min at 4°C (Hitachi CP100MX, Japan). The pellet was suspended in 10 ml pre-cooling PBS and centrifuged at 110,000 g for 70 min at 4°C again. Finally, the pellet was resuspended in pre-cooling 50 μL PBS and stored at −80°C until use.

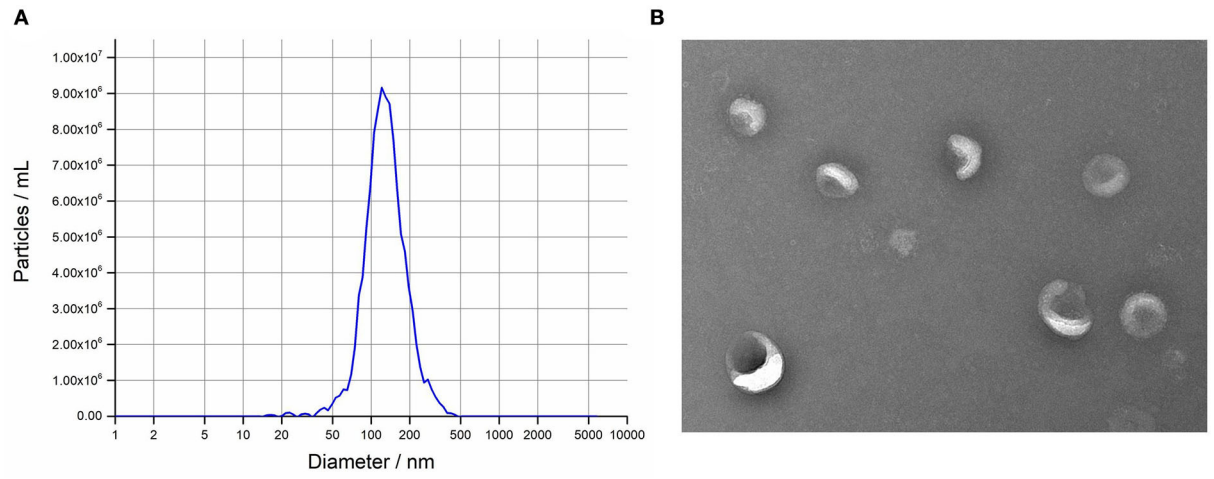


FIGURE 1
Isolation and characterization of intrauterine exosomes. **(A)** Transmission electron microscopy images showed that the isolated vesicles were oval or bowl-shaped. Scale bar, 200 nm. **(B)** Nanoparticle tracking analysis results demonstrated that the particle size distribution was in exosome-enriched fractions.

2.3. Nanoparticle tracking analysis and transmission electron microscopy

Nanoparticle tracking analysis (NTA) was performed for all samples by ZetaView S/N 17-310 (Particle metrix, Germany). One sample was randomly selected to detect the morphology using transmission electron microscopy (TEM). For this sample, a 10 μL aliquot was placed on a carbon-coated copper grid for 1 min and stained with a drop of 2% phosphotungstic acid for 1 min. After blotting off the excess liquid using filter paper, the grids were dried for 15 min at room temperature. Micrographs were then observed using a transmission electron microscope (Hitachi HT-7700, Japan).

2.4. Small RNA library construction

Total RNA in exosomes from uterine flushing fluids was extracted using the miRNeasy Serum/Plasma Kit (Qiagen, Germany) according to the manufacturer's instructions. The integrity and concentration were determined by the Agilent 2100 Bioanalyzer (Agilent Technology, USA). A total of 10 ng total RNA was used to construct the small RNA library using the TruSeq Small RNA Sample Prep Kits (Illumina, USA). In brief, adapters were ligated to both ends of the total RNA. Complementary DNA (cDNA) was then synthesized through reverse transcribe. Pieces between 147 and 157 bp were collected to construct a small RNA library. Agilent Technologies 2100 Bioanalyze was used to check the size and purity of the library. Finally, the library was sorted using the Illumina HiSeq X Ten platform. In the end, 150 bp paired-end reads were generated. Small RNA sequencing and analysis were performed by OE biotech (Shanghai, China). Data are available from the GEO database GSE216746.

2.5. Bioinformatics analysis

Clean reads were obtained by removing reads with low quality, 5' primer contaminants and poly (A). Reads without 3' adapter, insert tag, shorter than 15 nt and longer than 41 nt were also filtered. Q20 quality control was performed on the sequence to remove low quality reads where Q20 did not reach 80%. The length distribution of the clean reads in the reference genome (ARS-UCD 1.2) was determined. Non-coding RNAs consisting of rRNAs, tRNAs, small nuclear RNAs, and small nucleolar RNAs were aligned and then subjected to the BLAST search against Rfam v.10.1 (16). The known miRNAs were annotated by aligning against the miRBase v.21 database (17). Unannotated small RNAs were analyzed by miRdeep2 to predict novel miRNAs (18).

MiRNA expression levels were estimated by TPM (transcript per million). Differential expression analysis between two groups was conducted using the "DESeq" package in R (19). *P*-values were corrected for multiple testing using the Benjamini-Hochberg method (P_{BH}). The miRNAs with *p*-value < 0.05 and $|\log_2 \text{FoldChange}| > 1$ were considered as significantly differentially expressed miRNAs. Prediction of the target genes for the identified differentially expressed miRNAs was performed using the miRanda program (20) with the parameter of $S \geq 150$, $\Delta G \leq -30$ kcal/mol and demand strict 5' seed pairing. Gene Ontology (GO)

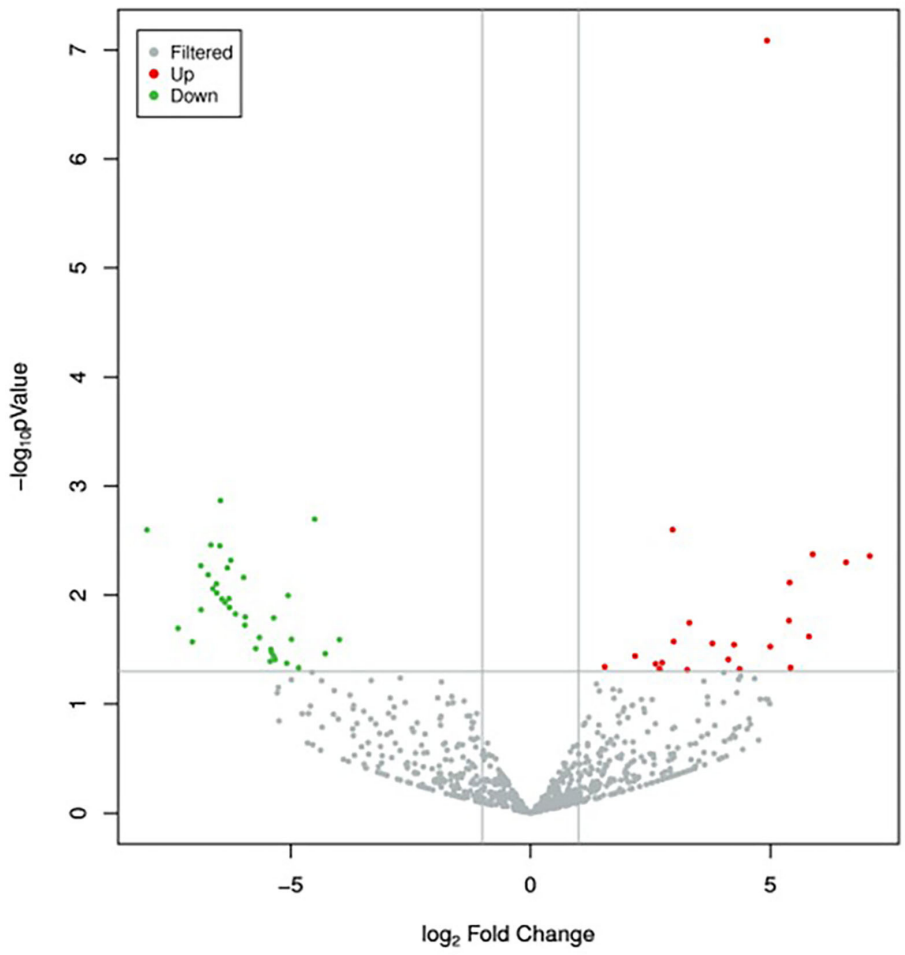


FIGURE 2
Volcano plot of differentially expressed miRNAs between the two groups. Red dots and green dots denote miRNAs were up-regulated and down-regulated in the group of donor cows, respectively. The significance cut-off was set to a *p*-value of 0.05 and $|\log_2 FoldChange|$of 1.

enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of these predicted target genes were performed using R based on the hypergeometric distribution (21). The false discovery rate of 0.05 and ListHits of five were set as the selection threshold for GO and KEGG analyses.

3. Results

3.1. Exosome isolation and characterization

The isolated exosomes from uterine flushing fluids were identified by NTA and TEM. The isolated vesicles have a classic exosome size. On average, the most widely distributed particle diameter of the six samples was 126.9 nm (Figure 1A). The result of TEM showed that the vesicles were bowl- shaped, which have the classic exosome morphology (Figure 1B).

3.2. Differential miRNA expression analysis

A total of 895 unique miRNAs were differentially expressed between the donor and recipient cows. Among them, 60 miRNAs were significantly differentially expressed (*p*-value<0.05, $|\log_2 FoldChange|>1$) (Figure 2). Twenty-two miRNAs were upregulated and 38 miRNAs downregulated in the group of donor cows (Table 2). The top 10 differentially expressed miRNAs were novel352_mature, bta-miR-1388-3p, bta-miR-211, novel548_mature, novel530_mature, bta-miR-1224, novel275_star, novel190_mature, novel170_mature, bta-miR-2904.

10.3389/fvets.2022.1078394

TABLE 2 Significantly differentially expressed miRNAs in intrauterine exosomes between the donor and recipient cows on 7 days after fertilization (donor group vs. recipient group).

miRNA_ID	|log2 FoldChang|	*P*-value	P_{BH}	Regulation
bta-miR-184	4.931	<0.001	<0.001	Up
bta-miR-660	2.964	0.003	0.382	Up
bta-miR-211	7.074	0.004	0.382	Up
bta-miR-2285bc	5.885	0.004	0.382	Up
bta-miR-2904	6.575	0.005	0.382	Up
novel399_mature	5.402	0.008	0.408	Up
novel490_mature	5.387	0.017	0.503	Up
bta-miR-129-5p	3.313	0.018	0.510	Up
novel527_mature	5.805	0.024	0.606	Up
bta-miR-10225b	2.987	0.027	0.606	Up
bta-miR-154c	3.791	0.028	0.610	Up
novel459_mature	4.244	0.029	0.614	Up
bta-miR-191b	4.998	0.030	0.621	Up
bta-miR-504	2.179	0.036	0.664	Up
bta-miR-365-5p	4.124	0.039	0.686	Up
bta-miR-2387	2.746	0.042	0.698	Up
novel288_mature	2.611	0.043	0.698	Up
bta-miR-2285q	5.418	0.046	0.711	Up
bta-miR-200b	1.549	0.046	0.711	Up
novel30_mature	4.359	0.048	0.711	Up
bta-miR-1306	3.269	0.048	0.711	Up
bta-miR-885	2.692	0.048	0.711	Up
novel439_mature	6.457	0.001	0.382	Down
novel142_mature	4.494	0.002	0.382	Down
novel352_mature	7.989	0.003	0.382	Down
novel190_mature	6.66	0.003	0.382	Down
novel145_mature	6.472	0.004	0.382	Down
novel530_mature	6.866	0.005	0.382	Down
novel159_mature	6.245	0.005	0.382	Down
novel504_mature	6.315	0.006	0.382	Down
novel275_star	6.714	0.007	0.405	Down
novel425_mature	5.975	0.007	0.405	Down
novel165_mature	6.545	0.008	0.408	Down
novel170_mature	6.614	0.009	0.427	Down
novel138_mature	6.536	0.010	0.430	Down
novel484_mature	5.049	0.010	0.430	Down
novel473_mature	6.427	0.011	0.430	Down
novel471_mature	6.283	0.011	0.430	Down
novel71_mature	6.365	0.012	0.430	Down
novel155_mature	6.361	0.012	0.430	Down
novel406_mature	6.272	0.013	0.457	Down
bta-miR-1224	6.863	0.014	0.462	Down
novel513_mature	6.144	0.015	0.485	Down
novel180_mature	5.941	0.016	0.491	Down
novel317_mature	5.35	0.016	0.491	Down
novel378_mature	5.949	0.019	0.519	Down
bta-miR-1388-3p	7.343	0.020	0.538	Down
novel577_mature	5.645	0.025	0.606	Down
novel318_mature	4.978	0.025	0.606	Down
novel311_mature	3.978	0.026	0.606	Down
novel548_mature	7.047	0.027	0.606	Down
novel150_mature	5.723	0.031	0.630	Down
novel259_mature	5.408	0.032	0.630	Down
novel408_mature	5.391	0.034	0.660	Down
novel186_mature	4.273	0.034	0.660	Down
novel110_mature	5.351	0.036	0.664	Down
novel526_mature	5.321	0.039	0.686	Down
novel119_mature	5.423	0.040	0.698	Down
bta-miR-1343-5p	5.077	0.042	0.698	Down
novel405_mature	4.828	0.046	0.711	Down

3.3. GO enrichment and KEGG pathway analyses of predicted target genes of the differentially expressed miRNAs

Target genes of the differentially expressed miRNAs were predicted using the miRanda program. Subsequently, GO enrichment analysis was performed to explore the potential functions of these target genes, which was used to infer the function of the differentially expressed miRNAs. The GO analysis results were assigned to biological processes (BP), cellular component (CC), and molecular function (MF).

Among the 60 differentially expressed miRNAs, 55 miRNAs were available to predict 9775 non-repeating target genes. 1568, 299, and 476 GO terms that were significant for BP, CC, and MF, respectively. Based on the top 30 GO terms (Figure 3), the GO analysis showed that the target genes were enriched in cell differentiation, cell adhesion, and cell junction. These processes play an important role in embryo implantation (22).

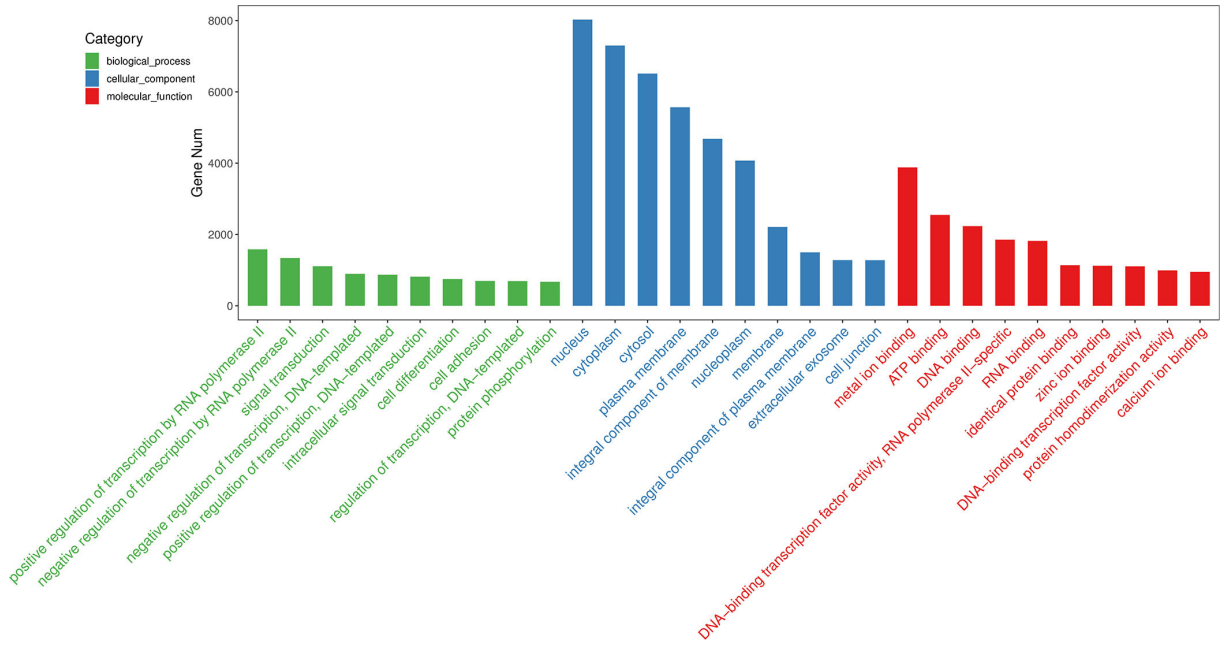


FIGURE 3
GO analysis of predicted target genes of differentially expressed miRNAs. Top 10 terms of biological process, cellular component, and molecular function were shown.

For KEGG analysis, eighty-three terms were significant. Some KEGG pathways in the top 20 terms, such as focal adhesion, adherens junction, tight junction, Rap1 signaling pathway and MAPK signaling pathway, were related to embryo implantation and endometrial development (23–27) (Figure 4).

4. Discussion

In the current study, miRNA expression profiles in intrauterine exosomes of donor and recipient cows on day 7 after fertilization were characterized. A total of 60 differentially expressed miRNAs were found. GO and KEGG analysis showed that the target genes of these miRNAs were mainly enriched in several biological processes or pathways associated with embryo implantation and endometrial development. These results demonstrated that the miRNAs in intrauterine exosomes could be significantly changed by the embryos on day 7 and might participate in the embryo implantation.

Embryo implantation is a complex process. The miRNA-mRNA regulatory system has been found to play an important role in the successful implantation of embryos (28). In this study, GO analysis showed that the top 30 biological processes were mainly enriched in cell differentiation, cell adhesion and cell junction, which are related to embryo implantation (29–32). For the KEGG analysis, the most significant term is Rap1 signaling pathway. The Rap1 signaling pathway has many important biological functions such as control of cell adhesion, cell junction, and cell proliferation (23, 33, 34). In addition, among the top 20 enriched pathways, some pathways such as MAPK signaling pathway, focal adhesion, adherens junction, and cell adhesion molecules provided more evidence that the differentially expressed miRNAs identified in this study were involved in the embryo implantation and endometrial development.

It has been reported that some miRNAs in the embryo- and endometrium-derived exosomes affect embryo implantation (15, 35, 36). A recent study, which has a similar research strategy as our study, found nine miRNAs in the UFs were differently expressed between inseminated and non-inseminated Japanese Black cows (37). Nevertheless, no overlapping miRNAs were found between this study and ours. This consequence might be due to the difference in the cattle breeds.

This study demonstrated that the presence of intrauterine exosomes, especially the miRNAs in intrauterine exosomes, played an important role in the early pregnancy in cattle. But the origin of exosomes was not determined. Due to the limitation of sample sources, the sample size of this experiment is small. In order to reduce the experimental error caused by large individual differences, we chose cows with similar weight and age as much as possible. In order to avoid the occurrence of false positives, we used the analysis method of multiple comparisons to analyze the experimental data. In spite of we selected donors and recipients with similar body weight and

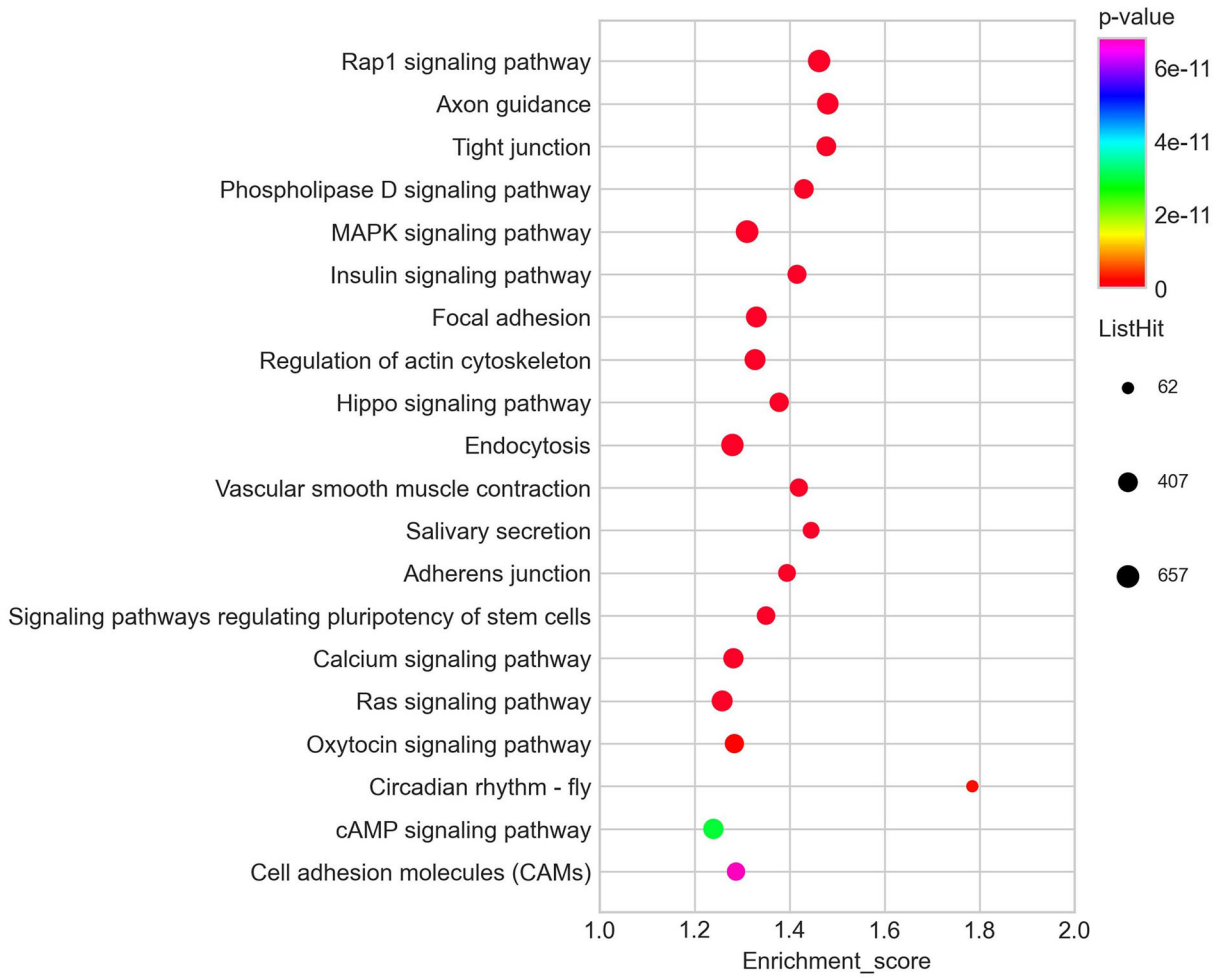


FIGURE 4
Top 20 KEGG pathways of predicted target genes of differentially expressed miRNAs.

age, but the influence of individual difference on the results cannot be ruled out, as well as a possible effect stemming from the low sample size used in this study. These issues lead to only one miRNA (bta-miR-184) being significantly differential expressed after multiple testing correction. The deletion of miR-184 in the process of egg cell expulsion and early embryonic development in drosophila melanogaster can lead to a variety of serious pathologies, which can lead to the loss of egg function. Targeted regulation of these genes by miR-184 may be necessary for oogenesis and early embryo development (38). MiR-184 has been reported to play roles in embryo implantation. In dairy goats, high throughput sequencing results of miRNA in the endometrial receptivity (RE) receptivity (PE) of dairy goats showed that the enrichment of miR-184 and RE was 31 times higher than that of PE ($P < 0.01$) (39), and miR-184 promotes the apoptosis of endometrial epithelial cells by downregulating STC2 *via* the RAS/RAF/MEK/ERK pathway and is involved in the establishment of receptive endometrium (40).

For the differentially expressed miRNAs (p-value < 0.05), although the function of them was not verified, the GO and KEGG analysis indicated that they might be involved in the process of embryo implantation in cattle. The impact of these miRNAs on the establishment of pregnancy in cattle warrants further investigations.

5. Conclusion

In conclusion, 60 miRNAs showing significant differences in expression were detected by comparing intrauterine exosomes of donor vs. recipient cows on 7 days after fertilization. GO functional enrichment analysis and KEGG signaling pathway analysis showed that these miRNAs were involved

in biological processes related to cell adhesion and focal adhesion. This strongly suggests that miRNAs in intrauterine exosomes exert some kind of influence on the early stages of cow pregnancy. Our study provides an experimental basis to screen and explore the intrauterine exosomal miRNA affecting embryo implantation.

Data availability statement

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found at: GEO, NCBI; GSE216746.

Ethics statement

The animal study was reviewed and approved by Committee for Experimental Animals at Henan Academy of Agricultural Sciences.

Author contributions

YZ: methodology, investigation, data curation, and writing–original draft. QS and FC: methodology, data curation, and formal analysis. QC: methodology, investigation, and data curation. YF, ZZ, and XQ: methodology and data curation. DA and GB: formal analysis and writing–review and editing. EW: conceptualization, funding acquisition, resources, supervision, and writing–review and editing. SL: conceptualization, methodology, investigation, data curation, formal analysis, writing–original draft, and writing–review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Science and Technology Program of Henan Province (222102110018), the Science-Technology Foundation for Outstanding Young Scientists of Henan Academy of Agricultural Sciences (2022YQ20), the Key Scientific and Technological Special Projects of Henan Province (221100110200), the Central Government Guides Local Science and Technology Development Fund Projects (Z20221343039), and the Program of the National Beef Cattle and Yak Industrial Technology System (CARS-37).

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

1. Wiltbank MC, Baez GM, Garcia-Guerra A, Toledo MZ, Monteiro PL, Melo LF, et al. Pivotal periods for pregnancy loss during the first trimester of gestation in lactating dairy cows. *Theriogenology.* (2016) 86:239–53. doi: 10.1016/j.theriogenology.2016.04.037

2. Liao XG Li YL, Gao RF, Geng YQ, Chen XM, Liu XQ, et al. Folate deficiency decreases apoptosis of endometrium decidual cells in pregnant mice *via* the mitochondrial pathway. *Nutrients.* (2015) 7:1916–32. doi: 10.3390/nu7031916

3. Silva F, da Silva GF, Vieira BS, Neto AL, Rocha CC, Lo Turco EG, et al. Peri-estrus ovarian, uterine, and hormonal variables determine the uterine luminal fluid metabolome in beef heifers. *Biol Reprod.* (2021) 105:1140–53. doi: 10.1093/biolre/ioab149

4. Kim GB, Shon O-J, Seo M-S, Choi Y, Park WT, Lee GW. Mesenchymal stem cell-derived exosomes and their therapeutic potential for osteoarthritis. *Biology.* (2021) 10:285. doi: 10.3390/biology10040285

5. Thery C. Exosomes: secreted vesicles and intercellular communications. *Biol Rep.* (2011) 3:15. doi: 10.3410/B3-15

6. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science.* (2020) 367:105247. doi: 10.1126/science.aau6977

7. O'Neil EV, Burns GW, Spencer TE. Extracellular vesicles: novel regulators of conceptus-uterine interactions? *Theriogenology.* (2020) 150:106–12. doi: 10.1016/j.theriogenology.2020.01.083

8. Capra E, Lange-Consiglio A. The biological function of extracellular vesicles during fertilization, early embryo-maternal crosstalk and their involvement in reproduction: review and overview. *Biomolecules.* (2020) 10:1510. doi: 10.3390/biom10111510

9. Burns G, Brooks K, Wildung M, Navakanitworakul R, Christenson LK, Spencer TE. Extracellular vesicles in luminal fluid of the ovine uterus. *PLoS ONE.* (2014) 9:e90913. doi: 10.1371/journal.pone.0090913

10. Kusama K, Nakamura K, Bai R, Nagaoka K, Sakurai T, Imakawa K. Intrauterine exosomes are required for bovine conceptus implantation. *Biochem Biophys Res Commun.* (2018) 495:1370–5. doi: 10.1016/j.bbrc.2017.11.176

11. Ruiz-Gonzalez I, Xu J, Wang X, Burghardt RC, Dunlap KA, Bazer FW. Exosomes, endogenous retroviruses and toll-like receptors: pregnancy recognition in ewes. *Reproduction.* (2015) 149:281–91. doi: 10.1530/REP-14-0538

12. Nakamura K, Kusama K, Bai R, Sakurai T, Isuzugawa K, Godkin JD, et al. Induction of ifnt-stimulated genes by conceptus-derived exosomes during the attachment period. *PLoS ONE.* (2016) 11:e0158278. doi: 10.1371/journal.pone.0158278

13. Czernek L, Düchler M. Exosomes as messengers between mother and fetus in pregnancy. *Int J Mol Sci.* (2020) 21:4264. doi: 10.3390/ijms21124264

14. Vilella F, Moreno-Moya JM, Balaguer N, Grasso A, Herrero M, Martínez S, et al. Hsa-Mir-30d, secreted by the human endometrium, is taken up by the pre-implantation embryo and might modify its transcriptome. *Development.* (2015) 142:3210–21. doi: 10.1242/dev.124289

15. Xie Y, Liu G, Zang X, Hu Q, Zhou C, Li Y, et al. Differential expression pattern of goat uterine fluids extracellular vesicles mirnas during peri-implantation. *Cells.* (2021) 10:2308. doi: 10.3390/cells10092308

16. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res.* (2003) 31:439–41. doi: 10.1093/nar/gkg006

17. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. Mirbase: tools for microrna genomics. *Nucleic Acids Res.* (2008) 36:D154–8. doi: 10.1093/nar/gkm952

18. Friedlander MR, Mackowiak SD Li N, Chen W, Rajewsky N. Mirdeep2 accurately identifies known and hundreds of novel microrna genes in seven animal clades. *Nucleic Acids Res.* (2012) 40:37–52. doi: 10.1093/nar/gkr688

19. Anders S. Analysing RNA-Seq data with the Deseq package. *Mol Biol.* (2010) 43:1–17.

20. 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

21. Kong D, Chen T, Zheng X, Yang T, Zhang Y, Shao J. Comparative profile of exosomal microrna in postmenopausal women with various bone mineral densities by small RNA sequencing. *Genomics.* (2021) 113:1514–21. doi: 10.1016/j.ygeno.2021.03.028

22. Ashary N, Tiwari A, Modi D. Embryo implantation: war in times of love. *Endocrinology.* (2018) 159:1188–98. doi: 10.1210/en.2017-03082

23. Gaonac'h-Lovejoy V, Boscher C, Delisle C, Gratton JP. Rap1 is involved in angiopoietin-1-induced cell-cell junction stabilization and endothelial cell sprouting. *Cells*. (2020) 9:155. doi: 10.3390/cells9010155

24. Albayrak IG, Azhari F, Colak EN, Balci BK, Ulgen E, Sezerman U, et al. Endometrial gene expression profiling of recurrent implantation failure after *in vitro* fertilization. *Mol Biol Rep.* (2021) 48:5075–82. doi: 10.1007/s11033-021-06502-x

25. Luan L, Ding T, Stinnett A, Reese J, Paria BC. Adherens junction proteins in the hamster uterus: their contributions to the success of implantation. *Biol Reprod.* (2011) 85:996–1004. doi: 10.1095/biolreprod.110.090126

26. Wang X, Matsumoto H, Zhao X, Das SK, Paria BC. Embryonic signals direct the formation of tight junctional permeability barrier in the decidualizing stroma during embryo implantation. *J Cell Sci.* (2004) 117:53–62. doi: 10.1242/jcs.00826

27. Zhang Y, Yang Z, Wu J. Signaling pathways and preimplantation development of mammalian embryos. *FEBS J.* (2007) 274:4349–59. doi: 10.1111/j.1742-4658.2007.05980.x

28. Salilew-Wondim D, Gebremedhn S, Hoelker M, Tholen E, Hailay T, Tesfaye D. The role of micrornas in mammalian fertility: from gametogenesis to embryo implantation. *Int J Mol Sci.* (2020) 21:585. doi: 10.3390/ijms21020585

29. Wu F, Mao D, Liu Y, Chen X, Xu H, Li TC, et al. Localization of Mucin 1 in endometrial luminal epithelium and its expression in women with reproductive failure during implantation window. *J Mol Histol.* (2019) 50:563–72. doi: 10.1007/s10735-019-09848-6

30. Poon CE, Madawala RJ, Dowland SN, Murphy CR. Nectin-3 is increased in the cell junctions of the uterine epithelium at implantation. *Reprod Sci.* (2016) 23:1580–92. doi: 10.1177/1933719116648216

31. Su R-W, Jia B, Ni H, Lei W, Yue S-L, Feng X-H, et al. Junctional adhesion molecule 2 mediates the interaction between hatched blastocyst and luminal epithelium: induction by progesterone and Lif. *PLoS ONE.* (2012) 7:e34325. doi: 10.1371/journal.pone.0034325

32. Ye X, Herr DR, Diao H, Rivera R, Chun J. Unique uterine localization and regulation may differentiate Lpa3 from other lysophospholipid receptors for its role in embryo implantation. *Fert Ster.* (2011) 95:2107–13. doi: 10.1016/j.fertnstert.2011.02.024

33. Boettner B, Van Aelst L. Control of cell adhesion dynamics by Rap1 signaling. *Curr Opin Cell Biol.* (2009) 21:684–93. doi: 10.1016/j.ceb.2009.06.004

34. Li Q, Teng Y, Wang J, Yu M, Li Y, Zheng H. Rap1 promotes proliferation and migration of vascular smooth muscle cell *via* the ERK pathway. *Pathol Res Pract.* (2018) 214:1045–50. doi: 10.1016/j.prp.2018.04.007

35. Kurian NK, Modi D. Extracellular vesicle mediated embryo-endometrial cross talk during implantation and in pregnancy. *J Assist Reprod Genet.* (2019) 36:189–98. doi: 10.1007/s10815-018-1343-x

36. Wang X, Li Q, Xie T, Yuan M, Sheng X, Qi X, et al. Exosomes from bovine endometrial epithelial cells ensure trophoblast cell development by MIR-218 targeting secreted frizzled related protein 2. *J Cell Physiol.* (2021) 236:4565–79. doi: 10.1002/jcp.30180

37. Kusama K, Rashid MB, Kowsar R, Marey MA, Talukder AK, Nagaoka K, et al. Day 7 embryos change the proteomics and exosomal micro-rnas content of bovine uterine fluid: involvement of innate immune functions. *Front Genet.* (2021) 12:676791. doi: 10.3389/fgene.2021.676791

38. Nicola I. Mir-184 Has Multiple roles in drosophila female germline development. *Dev Cell.* (2009) 17:123–33. doi: 10.1016/j.devcel.2009.06.008

39. Song Y, An X, Zhang L, Fu M, Peng J, Han P, et al. Identification and profiling of micrornas in goat endometrium during embryo implantation. *PLoS ONE.* (2015) 10:e0122202. doi: 10.1371/journal.pone.0122202

40. Cui J, Liu X, Yang L, Che S, Guo H, Han J, et al. Mir-184 combined with stc2 promotes endometrial epithelial cell apoptosis in dairy goats *via* Ras/Raf/Mek/Erk pathway. *Genes.* (2020) 11:1052. doi: 10.3390/genes11091052

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1065096

Check for updates

OPEN ACCESS

EDITED BY
Ibrar Muhammad Khan,
Fuyang Normal University, China

REVIEWED BY
Xiaoxue Zhang,
Gansu Agricultural University, China
Cai Yafei,
Nanjing Agricultural University, China
Rajwali Khan,
University of Agriculture, Peshawar, Pakistan

*CORRESPONDENCE
Bo Han,
bohan@cau.edu.cn

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

RECEIVED 09 October 2022
ACCEPTED 15 December 2022
PUBLISHED 04 January 2023

CITATION
Jia R, Xu L, Sun D and Han B (2023), Genetic marker identification of *SEC13* gene for milk production traits in Chinese holstein.
Front. Genet. 13:1065096.
doi: 10.3389/fgene.2022.1065096

COPYRIGHT
© 2023 Jia, Xu, Sun and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic marker identification of *SEC13* gene for milk production traits in Chinese holstein

Ruike Jia[1], Lingna Xu[1], Dongxiao Sun[1,2] and Bo Han[1]*

[1]Department of Animal Genetics and Breeding, College of Animal Science and Technology, Key Laboratory of Animal Genetics, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, National Engineering Laboratory for Animal Breeding, China Agricultural University, Beijing, China, [2]National Dairy Innovation Center, Hohhot, China

SEC13 homolog, nuclear pore and COPII coat complex component (*SEC13*) is the core component of the cytoplasmic COPII complex, which mediates material transport from the endoplasmic reticulum to the Golgi complex. Our preliminary work found that *SEC13* gene was differentially expressed in dairy cows during different stages of lactation, and involved in metabolic pathways of milk synthesis such as citric acid cycle, fatty acid, starch and sucrose metabolisms, so we considered that the *SEC13* might be a candidate gene affecting milk production traits. In this study, we detected the polymorphisms of *SEC13* gene and verified their genetic effects on milk yield and composition traits in a Chinese Holstein cow population. By sequencing the whole coding and partial flanking regions of *SEC13*, we found four single nucleotide polymorphisms (SNPs). Subsequent association analysis showed that these four SNPs were significantly associated with milk yield, fat yield, protein yield or protein percentage in the first and second lactations ($p \leq .0351$). We also found that two SNPs in *SEC13* formed one haplotype block by Haploview4.2, and the block was significantly associated with milk yield, fat yield, fat percentage, protein yield or protein percentage ($p \leq .0373$). In addition, we predicted the effect of SNP on 5′region on transcription factor binding sites (TFBSs), and found that the allele A of 22:g.54362761A>G could bind transcription factors (TFs) GATA5, GATA3, HOXD9, HOXA10, CDX1 and Hoxd13; and further dual-luciferase reporter assay verified that the allele A of this SNP inhibited the fluorescence activity. We speculate that the A allele of 22:g.54362761A>G might inhibit the transcriptional activity of *SEC13* gene by binding the TFs, which may be a cause mutation affecting the formation of milk production traits in dairy cows. In summary, we proved that *SEC13* has a significant genetic effect on milk production traits and the identified significant SNPs could be used as candidate genetic markers for GS SNP chips development; on the other hand, we verified the transcriptional regulation of 22:g.54362761A>G on *SEC13* gene, providing research direction for further function validation tests.

KEYWORDS

genetic marker, SNP, milk yield and composition, dairy cattle, dual-luciferase reporter assay

1 Introduction

The traits that affect the production efficiency of dairy cows are calving traits, milk production traits, longevity traits and so on (Weller and Ezra 2016; Zhang H. et al., 2021), among which, milk production traits are the most important economic traits in dairy cow breeding, including milk yield, fat yield, protein yield, fat percentage and protein percentage (Spelman et al., 1996). They are quantitative traits, controlled by multiple genes and easily affected by the environment. In addition, dairy cattle breeding is faced with problems such as long generation interval and slow progress (Wiggans et al., 2017). In 2009, developed countries began to widely use genomic selection (GS) for dairy cattle breeding. GS could shorten the generation gap, accelerate the population genetic progress, reduce the breeding costs, and almost double the genetic progress rate (Wiggans et al., 2017). Since the promotion and application of GS in China in 2012, the genetic progress of Chinese Holstein cows has been significantly improved. According to studies, adding functional gene information with greater genetic effects of target traits to single nucleotide polymorphism (SNP) marker data can improve the accuracy of genomic estimated breeding value prediction (Zhang et al., 2014; Zhang et al., 2015; de Las Heras-Saldana et al., 2020). Therefore, many researchers try to find the key genes/loci that have significant impacts on milk production traits through candidate gene analysis, genome-wide association analysis and omics data analysis strategies (Schennink et al., 2009; Canovas et al., 2013; Cui et al., 2014; Liang et al., 2017; Bhat et al., 2020). Studies have shown that the milk production traits of dairy cows can be significantly affected by SNPs in the genes (Jiang et al., 2010; Dux et al., 2018; Han et al., 2019a; Han et al., 2019b; Clancey et al., 2019; Jia et al., 2021; Du et al., 2022; Fu et al., 2022; Ye et al., 2022).

Previously, we used isobaric tag for relative and absolute quantification (iTRAQ) technique to study the proteomes of nine liver tissue samples from three Holstein cows during dry period, early and peak lactations, and found that *SEC13* homolog, nuclear pore and COPII coat complex component (*SEC13*) gene was differentially expressed among different lactations (dry period vs peak lactation: fold change = 1.37, p-value = .03134; early lactation vs. peak lactation: fold change = 1.22, p-value = .001002) and significantly enriched in the metabolic items and pathways related to milk synthesis, such as citric acid cycle, fatty acid, starch and sucrose metabolism, mTOR and PPAR signal pathways (Xu et al., 2019). *SEC13*, a member of the WD-Repeat protein family, is the core component of the COPII complex in the cytoplasm and nuclear pore complex on the nuclear membrane (Antonny and Schekman 2001; Stagg et al., 2006). Moreover, *SEC13* shuttles between the nucleus and the cytoplasm, acting in the ER-to-Golgi vesicular transport system (Enninga et al., 2003). *SEC13* interacts with *SEC31* to form the outer cage of the COPII vesicle coat needed for the transport of vesicle proteins from the endoplasmic reticulum (ER) to the Golgi matrix (Stagg et al., 2006; Stagg et al., 2008). COPII coat is used to transport newly synthesized proteins, including secretory and transmembrane proteins. As part of the COPII complex, *SEC13* is involved in the process of insulin stimulating glucose uptake by controlling the amount of glucose transporter 4 (GLUT4) in the plasma membrane (Martin et al., 2000; Karylowski et al., 2004; Huang and Czech 2007; Stockli et al., 2011). In addition, we found that an SNP (rs109645852; Chr.22:54428720) 67 kb away from *SEC13* gene (Chr.22:54326136.54361315; Cattle Quantitative Trait Locus Database; https://www.animalgenome.org/jbrowse/q?data=db%2FbovARS&loc=Chr.22%3A54243801.54530800&tracks=Milk%20composition%20-%20fat%2CGenbank%20annotations&highlight=Chr.22%3A54428718.54428722%20(%22Milk%20fat%20yield%20(daughter%20deviation)%22) was significantly associated with milk fat yield, which was identified in a reported genome-wide association study of Holstein cows (p = .0359) (Meredith et al., 2012). Therefore, we inferred that *SEC13* gene might be an important functional gene affecting milk production traits of dairy cows.

In this study, based on the candidate gene *SEC13* for milk production traits of dairy cows, the SNPs of this gene were detected, their genetic associations with milk yield, fat yield, fat percentage, protein yield and protein percentage were analyzed, and whether they could be used in the development of GS SNP chips was evaluated. In addition, we predicted the effect of identified SNPs on transcription factor binding site (TFBS), and verify the effect of SNP at 5′flanking region on the transcriptional activity of *SEC13* gene by dual-luciferase reporter experiment, thus speculating the causal mutation of milk production traits in dairy cows.

2 Materials and methods

2.1 Animal selection and phenotypic data collection

In this study, 947 cows from 45 Chinese Holstein sire families in Beijing were used as the experimental population. They were distributed in 22 farms of Beijing Shounong Animal Husbandry Development Co., Ltd. (Beijing, China), and raised under the same feeding conditions, with accurate pedigree and Dairy Herd Improvement (DHI) records. We used the phenotypic data of 947 first-lactation cows and 654 second-lactation cows (293 cows only completed the first lactation) for association analysis. The data of the entire lactation period of the parity of each cow was used as the individual milk yield phenotype. The actual total milk yield was multiplied by the corresponding estimated coefficient to get 305-day milk yield. The 305-day milk fat and protein content were calculated by multiplying the 305-day milk yield by the average percentage of milk fat and protein, respectively. The average milk fat and protein percentages were the ratio of total milk fat and protein contents to total milk yield, respectively.

Results of descriptive statistics for the milk yield and composition in the two lactations were shown in Supplementary Table S1. Frozen semen and cow blood samples were provided by Beijing Dairy Center (Beijing, China).

2.2 Genomic DNA extraction

We extracted genomic DNA from frozen semen of 45 bulls by salt-out procedure, and used a TIANamp Blood DNA Kit (Tiangen, Beijing, China) to extract DNA from the blood of 947 cows. We detected the quantity and quality of extracted DNA samples by a NanoDrop2000 spectrophotometer (Thermo Science, Hudson, NH, United States) and gel electrophoresis respectively.

2.3 SNP identification and estimation of linkage disequilibrium

We designed primers with Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) to amplify all the coding regions and the 2000 bp of 5′and 3′regions of *SEC13* gene. The primers were synthesized by Beijing Genomics Institute (BGI, Beijing, China). The genomic DNA of bull frozen semen was mixed with the same amount, then amplified by PCR (Supplementary Table S2). We used 2% gel electrophoresis to detect whether the PCR amplification products were qualified, and sequenced the qualified PCR amplification products by Sanger sequencing (BGI, Beijing, China). Then we compared the sequenced sequences with the reference sequences (ARS-UCD1.2) on NCBI-BLAST (https://blast.ncbi.nlm. nih. gov/Blast.cgi) to find the potential SNPs. Subsequently, we used Haploview4.2 (Broad Institute of MIT and Harvard, Cambridge, MA, United States) to estimate the degree of linkage disequilibrium (LD) between the identified SNPs. In addition, we genotyped 947 dairy cows using Genotyping by Target Sequencing (GBTS) technology in Boruidi Biotechnology Co., Ltd. (Hebei, China).

2.4 Association analyses on milk production traits

We used the SAS 9.4 software (SAS Institute Inc., Cary, NC, United States) to estimate the genetic associations of the SNPs or haplotype blocks with milk production traits, 305-day milk yield, fat yield, fat percentage, protein yield and protein percentage, on first or second lactation with the following animal model: $y = \mu + HYS + b \times M + G + a + e$; where y is the phenotypic value of each trait of each cow; μ is the overall mean; HYS is the fixed effect of farm (1–22: 22 farms), year (1–4: 2012–2015) and season (1: April-May; 2: June-August; 3: September-November and 4: December–March); M is the age of calving as a covariant, b is the regression coefficient of covariant M; G is the genotype or haplotype combination effect; a is the individual random additive genetic effect, the distribution is N (0, $\mathbf{A}\delta_a^2$), the additive genetic variance is δ_a^2; and e is random residual, the distribution is N (0, $\mathbf{I}\delta_e^2$), the unit matrix I and the residual variance δ_e^2. Bonferroni correction was carried out by multiple tests, the significance level was equal to the original *p*-value divided by the number of genotype or haplotype combinations. We also calculated the additive (a), dominant (d), and substitution (α) effects as follows: $a = \frac{AA-BB}{2}$;$d = AB - \frac{AA+BB}{2}$; $\alpha = a + d(q - p)$; where, AA, BB, and AB are the least square means of the milk production traits in the corresponding genotypes, p is the frequency of allele A, and q is the frequency of allele B.

2.5 Prediction and verification of SNP induced changes in gene transcription activity

We used Jaspar software (http://jaspar.genereg.net/) to predict whether SNP in the 5′flanking region of *SEC13* gene changed the transcription factor binding site (TFBS; relative score≥ .90).

Further, we used the dual-luciferase reporter assay to verify the effect of SNP site that affect the transcription factor binding on gene expression activity. For 22:g.54362761A>G, we synthesized the fragment with SNP site, A or G. The fragment that carried endonuclease sites KpnI and Nhel at ends, respectively, were cloned into pGL4.14 luciferase analysis vector (Promega, Madison, WI, United States). The constructed plasmid was sequenced to confirm the integrity of each insert. The Endo-free Plasmid Maxi Kit (Omega Bio-tek, Inc., Norcross, GA, United States) was used to extract plasmids needed for cell transfection. Human embryonic kidney (HEK) 293T cells were cultured in Dulbecco's modified Eagle's medium (Gibco; Thermo Fisher Scientific Inc., MA, United States) supplemented with 10% fetal bovine serum (FBS; Gibco) at 5% CO_2 and 37°C. The cells were seeded into 24-well plates with 2 × 105–107 cells per well before transfection. The cells were transiently transfected with liposome 2000 (Invitrogen; Thermo Fisher Scientific Inc.). For each well, 500 ng of the constructed plasmid was co-transfected along with 10 ng of pRL-TK Renilla luciferase reporter vector (Promega). 48 h after transfection, the cells were harvested and the luciferase activity was detected by a Dual-Luciferase Reporter Assay System (Promega). The relative fluorescence activity was calculated by the fluorescence activity ratio of firefly and renilla.

3 Results

3.1 SNPs identification

We totally found four SNPs in *SEC13* gene, one SNP, 22:g.54362761A>G, is located in 5′flanking region, 22:g.54334911G>A in intron and two SNPs, 22:g.54326411T>C

TABLE 1 Detailed information about SNPs identified in SEC13 gene and their genotypic and allelic frequencies. UTR: untranslated region.

SNP name	Rs-ID	Region	Genotype	Genotypic frequency	Allele	Allelic frequency
22:g.54362761A>G	rs135591064	5′flanking region	AA	.623	A	.7856
			AG	.3252	G	.2144
			GG	.0517		
22:g.54334911G>A	rs43599316	intron	AA	.2313	A	.4725
			AG	.4826	G	.5275
			GG	.2862		
22:g.54326411T>C	rs133320599	3′UTR	CC	.7645	C	.8786
			CT	.2281	T	.1214
			TT	.0074		
22:g.54326366C>T	rs208189354	3′UTR	CC	.0591	C	.2592
			CT	.4002	T	.7408
			TT	.5407		

and 22:g.54326366C>T, in 3′untranslated region (UTR) (Table 1). Additionally, the genotypic and allelic frequencies of all the identified SNPs were summarized in Table 1.

3.2 Association analysis between SNP and the five milk production traits

We analyzed the associations between the four SNPs and five milk production traits in dairy cows, including 305-day milk yield, fat yield, protein yield, fat percentage and protein percentage (Table 2). The SNP 22:g.54362761A>G was significantly associated with protein yield (p = .036) in the first lactation, milk yield (p = .0004), fat yield (p < .0001) and protein yield (p < .0001) in the second lactation. 22:g.54334911G>A was significantly associated with fat yield (p = .0351) in the first lactation, milk yield (p = .0033), fat yield (p < .0001), protein yield (p = .0001) and protein percentage (p = .0274) in the second lactation. 22:g.54326411T>C was significantly associated with protein yield (p = .0296) in the second lactation. 22:g.54326366C>T was significantly associated with milk yield (p = .0173), fat yield (p = .0008) and protein yield (p = .0028) in the first lactation. Accordingly, the results of additive, dominant and substitution effects for the four SNPs were shown in Supplementary Table S3.

3.3 Association between haplotype block and the five milk traits

We estimated the degree of LD among four identified SNPs using Haploview4.2, and inferred one haplotype block, including two SNPs, 22:g.54326366C>T and 22:g.54326411T>C, (D' = 1; Figure 1). In the block, the frequencies of H1 (TC), H2 (CC) and H3 (CT) haplotypes were 74.1%, 13.8% and 12.1%, respectively. The haplotype block of *SEC13* was significantly associated with milk yield (p = .046), fat yield (p = .0001), fat percentage (p = .0113), protein yield (p = .0056) and protein percentage (p = .0373) in the first lactation (Supplementary Table S4).

3.4 Effect of 22:g.54362761A>G on gene transcriptional activity

We predicted the TFBS changes for the SNP, 22:g.54362761A>G, in the 5′flanking region of *SEC13* gene by Jaspar software (relative score (RS) ≥.90). The result showed that allele A of 22:g.54362761A>G created binding sites for transcription factors (TFs) GATA5 (RS = .94), GATA3 (RS = .93), HOXD9 (RS = .93), HOXA10 (RS = .91), HOXD13 (RS = .90) and CDX1 (RS = .91; Table 3).

To further determine whether this SNP changed the transcription activity of *SEC13* gene, we constructed reporter plasmids containing two alleles A and G, respectively (Figure 2A). As shown in Figure 2B, the luciferase activities of the two recombinant plasmids were significantly higher than that of the empty vector (PGL4.14 + TK) and the blank cell controls (p < .01), which confirmed that the inserted fragment have transcriptional regulation function. The luciferase activity of G allele (PGL4.14(G)+TK) was significantly higher than that of A allele (PGL4.14(A)+TK; p < .01), suggesting that the 22:g.54362761A might inhibit the transcription activity of *SEC13* gene.

TABLE 2 Associations of four SNPs with milk production traits in Chinese Holstein cattle during two lactations.

SNPs	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
22: g.54362761A>G	1	AA (590)	10296 ± 62.3329	343.61 ± 2.7953	3.3501 ± .02578	304.34 ± 2.0345a	2.9577 ± .008228
		AG (308)	10389 ± 70.5347	342.64 ± 3.1043	3.3082 ± .02899	308.47 ± 2.2602b	2.968 ± .009488
		GG (49)	10329 ± 121.39	346.76 ± 5.0561	3.3665 ± .04899	305.25 ± 3.6856ab	2.9563 ± .01719
		P	.2343	.6475	.1119	.036	.3926
	2	AA (412)	10715 ± 63.3167Aa	384.74 ± 2.8456Aa	3.6115 ± .0262	316.65 ± 2.0709A	2.9664 ± .008361
		AG (212)	10534 ± 78.0057b	377.71 ± 3.3976Ab	3.6095 ± .03195	309.52 ± 2.4743Ba	2.9603 ± .01061
		GG (30)	10195 ± 152.38Bb	357.97 ± 6.2924B	3.53 ± .06133	297 ± 4.5878Bb	2.9419 ± .02173
		P	.0004	<.0001	.3769	<.0001	.4886
22: g.54334911G>A	1	AA (219)	10257 ± 74.9765	344.44 ± 3.2661ab	3.3598 ± .0307	303.45 ± 2.3785	2.9567 ± .01022
		AG (457)	10302 ± 65.3486	340.75 ± 2.9116a	3.3204 ± .02697	304.93 ± 2.1194	2.962 ± .00871
		GG (271)	10419 ± 72.0751	346.68 ± 3.1562b	3.3477 ± .02957	308.36 ± 2.2982	2.9621 ± .009768
		P	.0621	.0351	.222	.0582	.8314
	2	AA (145)	10451 ± 86.4939Aa	374.52 ± 3.7205Aa	3.6043 ± .03527	307.98 ± 2.7101Aa	2.9625 ± .01191ab
		AG (327)	10644 ± 68.659b	378.51 ± 3.0478Aa	3.5789 ± .0283	312.54 ± 2.2187Aa	2.9519 ± .009178ab
		GG (182)	10772 ± 76.8091Bb	390.83 ± 3.344B	3.6486 ± .03145	319.71 ± 2.4352B	2.981 ± .01047b
		P	.0033	<.0001	.0651	.0001	.0274
22: g.54326411T>C	1	CC (724)	10332 ± 61.5915	342.78 ± 2.7712	3.3323 ± .02551	305.76 ± 2.0168	2.9608 ± .008119
		CT (216)	10289 ± 75.681	345.21 ± 3.2968	3.3589 ± .03099	304.8 ± 2.4008	2.9624 ± .01032
		TT (7)	10803 ± 281.57	356.33 ± 11.4228	3.287 ± .1127	314.18 ± 8.3315	2.8937 ± .04078
		P	.1582	.3467	.4835	.4734	.2365
	2	CC (504)	10624 ± 61.7697	381.29 ± 2.789	3.6076 ± .02561	313.01 ± 2.0296a	2.9623 ± .008109
		CT (145)	10751 ± 86.3478	385.02 ± 3.721	3.6128 ± .03523	318.76 ± 2.7104b	2.9715 ± .01188
		TT (5)	10989 ± 341.35	387.11 ± 13.8933	3.5485 ± .1367	324 ± 10.1328ab	2.9434 ± .04927
		P	.1661	.4657	.8873	.0296	.628
22: g.54326366C>T	1	CC (56)	10417 ± 112.99ab	346.02 ± 4.7173ab	3.3286 ± .04562	306.98 ± 3.4384ab	2.9536 ± .01599
		CT (379)	10236 ± 67.6416a	338.53 ± 2.9983Aa	3.3209 ± .02786	302.32 ± 2.1827Aa	2.955 ± .00905

(Continued on following page)

TABLE 2 (*Continued*) Associations of four SNPs with milk production traits in Chinese Holstein cattle during two lactations.

SNPs	Lactation	Genotype (No.)	Milk yield (kg)	Fat yield (kg)	Fat percentage (%)	Protein yield (kg)	Protein percentage (%)
		TT (512)	10378 ± 64.3329b	346.59 ± 2.8714Bb	3.3522 ± .02657	307.69 ± 2.09Bb	2.9657 ± .008536
		P	.0173	.0008	.3494	.0028	.3602
	2	CC (36)	10807 ± 137.95	381.62 ± 5.7148	3.566 ± .05557	319.77 ± 4.1663	2.9653 ± .01965
		CT (262)	10695 ± 72.6125	382.4 ± 3.2	3.5963 ± .02985	316.3 ± 2.3298	2.9671 ± .009777
		TT (356)	10608 ± 66.2633	381.76 ± 2.9521	3.619 ± .02734	312.29 ± 2.1488	2.9617 ± .008821
		P	.2174	.9696	.5142	.0518	.8565

Note: The number in the table represents the mean ± standard deviation; the number in the bracket represents the number of cows for the corresponding genotype; *p*-value shows the significance for the genetic effects of SNPs; a, b, c within the same column with different superscripts means $p < .05$; and A, B, C within the same column with different superscripts means $p < .01$.

FIGURE 1
Linkage disequilibrium estimated between SNPs in SEC13 gene. The bright red boxes without numbers indicate complete LD (D′ = 1).

4 Discussion

Our previous study considered *SEC13* gene to be a candidate to affect milk production traits in dairy cows. In this study, we detected the polymorphisms of *SEC13* gene, and found that there was a significant genetic association between the SNP/haplotype block of the gene and milk production traits. Studies have shown that in the genomic relationship matrices, SNPs can be given different weights to make more accurate and less biased prediction of traits (Tiezzi and Maltecca 2015). Therefore, when adding the SNPs that have significant effects on milk production traits in this study to the commercial SNP chip, we can give different weights to these SNPs according to their effects to improve the accuracy of GS.

Coding sequence (CDS) is a key area for creature survival, many mutations in this region can lead to death, and DNA repair in this region is more efficient than other regions. A studies had shown that the number of SNP in the exon region is less than that in other regions (Frigola et al., 2017), which may be the reason why we didn't find SNP in the exon region. Transcription factors (TFs) can affect gene expression by combining to TFBSs to regulate the transcription of target genes (Calkhoven and Ab 1996). The SNP located at TFBS may affect the binding of TFs, resulting in differences in gene expression among individuals with different genotypes (Kasowski et al., 2010; McDaniell et al., 2010). In this study, we found that the mutation from allele A to G of 22: g.54362761A>G led to the disappearance of TFBSs for TFs GATA5, GATA3, HOXD9, HOXA10, HOXD13 and CDX1.

TABLE 3 Transcription factor binding sites (TFBSs) prediction for SNP in *SEC13* gene.

SNP name	Allele	Transcription factor	Relative score (RS ≥ .90)	Predicted binding site sequence
22:g.54362761A>G	A	GATA5	.94	CGATAAAA
		GATA3	.93	CGATAAAA
		HOXD9	.93	GCGATAAAAT
		HOXA10	.91	AGCGATAAAAT
		CDX1	.91	GCGATAAAA
		Hoxd13	.90	AGCGATAAAA
	G			

Note: The underlined nucleotides represent the position of the SNP.

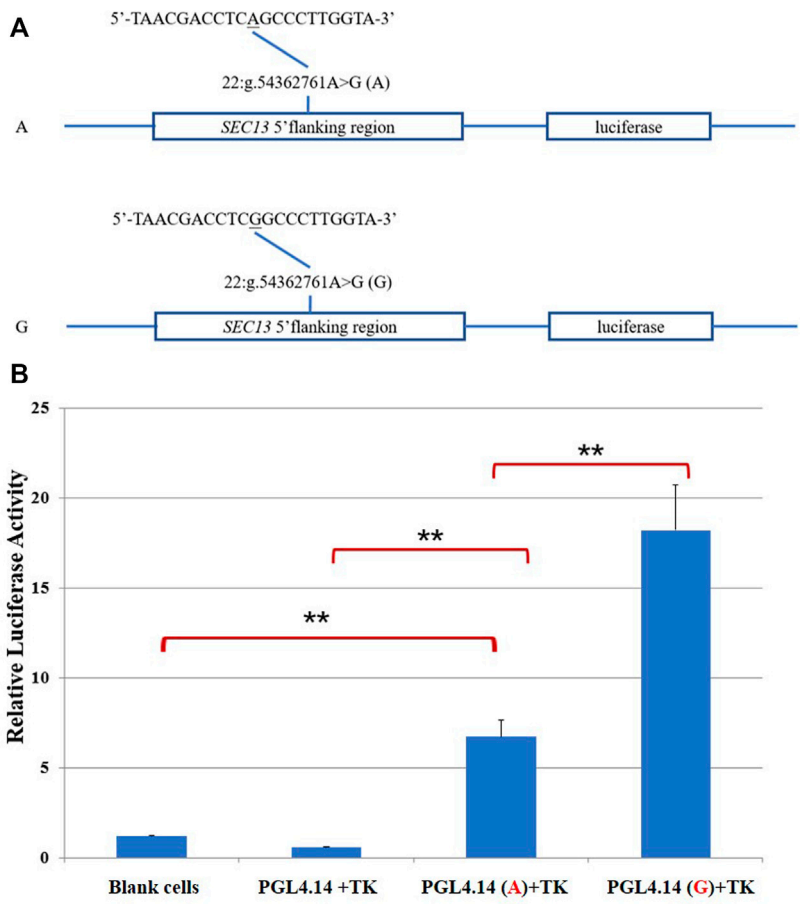


FIGURE 2
Dual-luciferase activity assay. **(A)**Sketches of recombinant plasmids with 22:g.54362761A>G in SEC13 gene. The underlined nucleotide was the SNP. **(B)** Luciferase activity analysis of the recombinant plasmids in HEK 293T cells. $^{**}p < .01$.

GATA family regulates cell reprogramming to induce stem cell differentiation and normal function of cells (Shu et al., 2015). GATA5 is a member of GATA family and plays an important role in cardiovascular disease (Shi et al., 2014; Messaoudi et al., 2015). Furthermore, GATA5 is involved in suppressing expression of the reprogramming genes and stemness markers in hepatocellular carcinoma cells (Feng et al., 2020). GATA3 is estimated to be the most highly

expressed transcription factor in the differentiated luminal epithelial cells lining the breast ductal structures and plays an important role in mammary gland development (Kouros-Mehr et al., 2006; Kouros-Mehr and Werb 2006). Moreover, GATA3 transcriptionally inhibits Slug expression, thereby inhibits cancer cell proliferation, migration and invasion (Zhang Z. et al., 2021). HOX is a family of many TFs that involved in embryonic development and the maintenance of normal tissues (Mallo and Alonso 2013). HOXD9 is closely related to many kinds of tumors. It can promote tumorigenesis and metastasis by increasing the expression level of other genes in gastric and breast cancer (Zhu et al., 2019; Xiong et al., 2020; Hu et al., 2022). Both HOXA10 and HOXD13 can promote the occurrence and development of various cancers by increasing the expression of genes and activating signal pathways (Chen et al., 2019; Cui et al., 2020; Yin and Guo 2021), nevertheless, both of them can inhibit the expression of genes and suppress the occurrence of prostate cancer (Hatanaka et al., 2019; Xu et al., 2021). CDX1 controls intestinal cell differentiation in the colon and has been shown to directly promote the expression of structural proteins important for epithelial differentiation, including cytokeratin 20 and villin (Chan et al., 2009; Arango et al., 2012). These TFs can activate or inhibit gene expression, and they may interact with each other to promote or inhibit the expression of *SEC13* gene. It has been reported that the *SEC13* gene is close to the quantitative trait locus (QTL) which has been found to have a significant effect on milk fat yield (Meredith et al., 2012). Another study showed that the QTL from the same region showed significant genetic effects on milk protein and fat traits (Bagnato et al., 2008). According to the phenotypic data of milk production traits of different genotypes, it was found that the milk, fat and protein yields of genotype AA was significantly higher than that of genotype GG of 22:g.54362761A>G. Further, the results of dual-luciferase reporter assay showed that when the A mutated to G in 22:g.54362761A>G, the transcriptional activity of *SEC13* gene increased significantly. Therefore, we speculated that the change of *SEC13* gene expression caused by SNP 22:g.54362761A>G may be one of the reasons for the phenotypic change of milk production traits in dairy cows. The specific mechanism of *SEC13* gene on the formation of milk production traits needs further experimental verification. SEC23B, which is also the core component of COPII complex with *SEC13*, mediates the transport of substances from the endoplasmic reticulum to the Golgi complex and plays an important role in professional secretory tissues/cells (Tao et al., 2012). This also provides an important clue for our in-depth research. Further, we can use secretory cell models such as bovine mammary epithelial cells to verify the influence of *SEC13* on the formation of milk production traits.

5 Conclusion

This study confirmed the genetic effects of *SEC13* on milk production traits of Chinese Holstein cows. The SNP 22:g.54362761A>G might be a causal mutation site for milk production traits, which might regulate the transcriptional activity of *SEC13* gene by binding transcription factors, the specific mechanism remains to be further verified. This study laid a foundation for the further functional verification of *SEC13* for milk synthesis in dairy cattle, and the significant SNP sites could be used as genetic markers for dairy cattle GS breeding.

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 The study was conducted in accordance with Guide for the Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University (Beijing, China; permit number: DK996).

Author contributions

BH and DS: Conceptualization, methodology and funding acquisition. RJ and LX: formal analysis. RJ: Visualization. RJ: Writing–original draft preparation. RJ and BH: Writing, review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by Inner Mongolia and Hohhot Science & Technology Plan (No.2020-Ke Ji Xing Meng-National Innovation Center-12; No.2021-National Dairy Innovation Center-3), National Natural Science Foundation of China (32072716; 31872330), Key R & D project of Ningxia Hui Autonomous Region (2021BEF02018),

and the Program for Changjiang Scholar and Innovation Research Team in University (IRT_15R62).

Acknowledgments

We appreciate Beijing Dairy Cattle Center for providing the semen and blood samples and phenotypic data.

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.2022.1065096/full#supplementary-material

SUPPLEMENTARY TABLE S1
Descriptive statistics of phenotypic values for milk production in two lactations.

SUPPLEMENTARY TABLE S2
Primers and procedures for PCR used in SNPs identification of *SEC13* gene.

SUPPLEMENTARY TABLE S3
Additive, dominant and allele substitution effects of four SNPs in *SEC13* gene on milk yield and composition traits in Chinese Holstein cows during two lactations.

SUPPLEMENTARY TABLE S4
Haplotype analysis for *SEC13* gene.

References

Antonny, B., and Schekman, R. (2001). ER export: Public transportation by the COPII coach. *Curr. Opin. Cell. Biol.* 13, 438–443.

Arango, D., Al-Obaidi, S., Williams, D. S., Dopeso, H., Mazzolini, R., Corner, G., et al. (2012). Villin expression is frequently lost in poorly differentiated colon cancer. *Am. J. Pathol.* 180, 1509–1521.

Bagnato, A., Schiavini, F., Rossoni, A., Maltecca, C., Dolezal, M., Medugorac, I., et al. (2008). Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. *J. Dairy Sci.* 91, 767–783. doi:10.3168/jds.2007-0507

Bhat, S. A., Ahmad, S. M., Ibeagha-Awemu, E. M., Mobashir, M., Dar, M. A., Mumtaz, P. T., et al. (2020). Comparative milk proteome analysis of Kashmiri and Jersey cattle identifies differential expression of key proteins involved in immune system regulation and milk quality. *BMC Genomics* 21, 161. doi:10.1186/s12864-020-6574-4

Calkhoven, C. F., and Ab, G. (1996). Multiple steps in the regulation of transcription-factor level and activity. *Biochem. J.* 317 (2), 329–342. doi:10.1042/bj3170329

Canovas, A., Rincon, G., Islas-Trejo, A., Jimenez-Flores, R., Laubscher, A., Medrano, J. F., et al. (2013). RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. *J. Dairy Sci.* 96, 2637–2648. doi:10.3168/jds.2012-6213

Chan, C. W., Wong, N. A., Liu, Y., Bicknell, D., Turley, H., Hollins, L., et al. (2009). Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1936–1941. doi:10.1073/pnas.0812904106

Chen, W., Wu, G., Zhu, Y., Zhang, W., Zhang, H., Zhou, Y., et al. (2019). HOXA10 deteriorates gastric cancer through activating JAK1/STAT3 signaling pathway. *Cancer Manag. Res.* 11, 6625–6635. doi:10.2147/CMAR.S201342

Clancey, E., Kiser, J. N., Moraes, J. G. N., Dalton, J. C., Spencer, T. E., and Neibergs, H. L. (2019). Genome-wide association analysis and gene set enrichment analysis with SNP data identify genes associated with 305-day milk yield in Holstein dairy cows. *Anim. Genet.* 50, 254–258. doi:10.1111/age.12792

Cui, X. G., Hou, Y. L., Yang, S. H., Xie, Y., Zhang, S. L., Zhang, Y., et al. (2014). Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing. *BMC Genomics* 15, 226. doi:10.1186/1471-2164-15-226

Cui, Y. P., Xie, M., Pan, W. X., Zhang, Z. Y., and Li, W. F. (2020). HOXA10 promotes the development of bladder cancer through regulating FOSL1. *Eur. Rev. Med. Pharmacol. Sci.* 24, 2945–2954. doi:10.26355/eurrev_202003_20659

de Las Heras-Saldana, S., Lopez, B. I., Moghaddar, N., Park, W., Park, J. E., Chung, K. Y., et al. (2020). Use of gene expression and whole-genome sequence information to improve the accuracy of genomic prediction for carcass traits in Hanwoo cattle. *Genet. Sel. Evol.* 52, 54. doi:10.1186/s12711-020-00574-2

Du, A., Zhao, F., Liu, Y., Xu, L., Chen, K., Sun, D., et al. (2022). Genetic polymorphisms of PKLR gene and their associations with milk production traits in Chinese Holstein cows. *Front. Genet.* 13, 1002706. doi:10.3389/fgene.2022.1002706

Dux, M., Muranowicz, M., Siadkowska, E., Robakowska-Hyzorek, D., Flisikowski, K., Bagnicka, E., et al. (2018). Association of SNP and STR polymorphisms of insulin-like growth factor 2 receptor (IGF2R) gene with milk traits in Holstein-Friesian cows. *J. Dairy Res.* 85, 138–141. doi:10.1017/S0022029918000110

Enninga, J., Levay, A., and Fontoura, B. M. (2003). Sec13 shuttles between the nucleus and the cytoplasm and stably interacts with Nup96 at the nuclear pore complex. *Mol. Cell. Biol.* 23, 7271–7284. doi:10.1128/mcb.23.20.7271-7284.2003

Feng, H., Lin, B., Zheng, Y., Xu, J., Zhou, Y., Liu, K., et al. (2020). Overexpression of GATA5 stimulates paclitaxel to inhibit malignant behaviors of hepatocellular carcinoma cells. *Cell. J.* 22, 89–100. doi:10.22074/cellj.2020.6894

Frigola, J., Sabarinathan, R., Mularoni, L., Muinos, F., Gonzalez-Perez, A., and Lopez-Bigas, N. (2017). Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* 49, 1684–1692. doi:10.1038/ng.3991

Fu, Y., Jia, R., Xu, L., Su, D., Li, Y., Liu, L., et al. (2022). Fatty acid desaturase 2 affects the milk-production traits in Chinese Holsteins. *Anim. Genet.* 53, 422–426. doi:10.1111/age.13192

Han, B., Yuan, Y., Liang, R., Li, Y., Liu, L., and Sun, D. (2019a). Genetic effects of LPIN1 polymorphisms on milk production traits in dairy cattle. *Genes (Basel)* 10, 265. doi:10.3390/genes10040265

Han, B., Yuan, Y., Shi, L., Li, Y., Liu, L., and Sun, D. (2019b). Identification of single nucleotide polymorphisms of PIK3R1 and DUSP1 genes and their genetic associations with milk production traits in dairy cows. *J. Anim. Sci. Biotechnol.* 10, 81. doi:10.1186/s40104-019-0392-z

Hatanaka, Y., de Velasco, M. A., Oki, T., Shimizu, N., Nozawa, M., Yoshimura, K., et al. (2019). HOXA10 expression profiling in prostate cancer. *Prostate* 79, 554–563. doi:10.1002/pros.23761

Hu, X. C., Chu, J., Zhou, Y., Li, C. C., Zhou, G. J., and Jiang, G. Q. (2022). HOXD9 transcriptionally induced UXT facilitate breast cancer progression via

epigenetic modification of RND3. *Cell. Signal* 90, 110188. doi:10.1016/j.cellsig.2021.110188

Huang, S., and Czech, M. P. (2007). The GLUT4 glucose transporter. *Cell. Metab.* 5, 237–252. doi:10.1016/j.cmet.2007.03.006

Jia, R., Fu, Y., Xu, L., Li, H., Li, Y., Liu, L., et al. (2021). Associations between polymorphisms of SLC22A7, NGFR, ARNTL and PPP2R2B genes and Milk production traits in Chinese Holstein. *BMC Genom Data* 22, 47. doi:10.1186/s12863-021-01002-0

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., et al. (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One* 5, e13661. doi:10.1371/journal.pone.0013661

Karylowski, O., Zeigerer, A., Cohen, A., and McGraw, T. E. (2004). GLUT4 is retained by an intracellular cycle of vesicle formation and fusion with endosomes. *Mol. Biol. Cell.* 15, 870–882. doi:10.1091/mbc.e03-07-0517

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., et al. (2010). Variation in transcription factor binding among humans. *Science* 328, 232–235. doi:10.1126/science.1183621

Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D., and Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell.* 127, 1041–1055. doi:10.1016/j.cell.2006.09.048

Kouros-Mehr, H., and Werb, Z. (2006). Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. *Dev. Dyn.* 235, 3404–3412. doi:10.1002/dvdy.20978

Liang, R., Han, B., Li, Q., Yuan, Y., Li, J., and Sun, D. (2017). Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Sci. Rep.* 7, 6396. doi:10.1038/s41598-017-06634-w

Mallo, M., and Alonso, C. R. (2013). The regulation of Hox gene expression during animal development. *Development* 140, 3951–3963. doi:10.1242/dev.068346

Martin, S., Millar, C. A., Lyttle, C. T., Meerloo, T., Marsh, B. J., Gould, G. W., et al. (2000). Effects of insulin on intracellular GLUT4 vesicles in adipocytes: Evidence for a secretory mode of regulation. *J. Cell. Sci.* 113 Pt 19, 3427–3438. doi:10.1242/jcs.113.19.3427

McDaniell, R., Lee, B. K., Song, L., Liu, Z., Boyle, A. P., Erdos, M. R., et al. (2010). Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* 328, 235–239. doi:10.1126/science.1184655

Meredith, B. K., Kearney, F. J., Finlay, E. K., Bradley, D. G., Fahey, A. G., Berry, D. P., et al. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet.* 13, 21. doi:10.1186/1471-2156-13-21

Messaoudi, S., He, Y., Gutsol, A., Wight, A., Hebert, R. L., Vilmundarson, R. O., et al. (2015). Endothelial Gata5 transcription factor regulates blood pressure. *Nat. Commun.* 6, 8835. doi:10.1038/ncomms9835

Schennink, A., Stoop, W. M., Visker, M. H., van der Poel, J. J., Bovenhuis, H., and van Arendonk, J. A. (2009). Short communication: Genome-wide scan for bovine milk-fat composition. II. Quantitative trait loci for long-chain fatty acids. *J. Dairy Sci.* 92, 4676–4682. doi:10.3168/jds.2008-1965

Shi, L. M., Tao, J. W., Qiu, X. B., Wang, J., Yuan, F., Xu, L., et al. (2014). GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. *Int. J. Mol. Med.* 33, 1219–1226. doi:10.3892/ijmm.2014.1700

Shu, J., Zhang, K., Zhang, M., Yao, A., Shao, S., Du, F., et al. (2015). GATA family members as inducers for cellular reprogramming to pluripotency. *Cell. Res.* 25, 169–180. doi:10.1038/cr.2015.6

Spelman, R. J., Coppieters, W., Karim, L., van Arendonk, J. A., and Bovenhuis, H. (1996). Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144, 1799–1808. doi:10.1093/genetics/144.4.1799

Stagg, S. M., Gurkan, C., Fowler, D. M., LaPointe, P., Foss, T. R., Potter, C. S., et al. (2006). Structure of the Sec13/31 COPII coat cage. *Nature* 439, 234–238. doi:10.1038/nature04339

Stagg, S. M., LaPointe, P., Razvi, A., Gurkan, C., Potter, C. S., Carragher, B., et al. (2008). Structural basis for cargo regulation of COPII coat assembly. *Cell.* 134, 474–484. doi:10.1016/j.cell.2008.06.024

Stockli, J., Fazakerley, D. J., and James, D. E. (2011). GLUT4 exocytosis. *J. Cell. Sci.* 124, 4147–4159. doi:10.1242/jcs.097063

Tao, J., Zhu, M., Wang, H., Afelik, S., Vasievich, M. P., Chen, X. W., et al. (2012). SEC23B is required for the maintenance of murine professional secretory tissues. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2001–E2009. doi:10.1073/pnas.1209207109

Tiezzi, F., and Maltecca, C. (2015). Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. *Genet. Sel. Evol.* 47, 24. doi:10.1186/s12711-015-0100-1

Weller, J. I., and Ezra, E. (2016). Genetic analysis of calving traits by the multi-trait individual animal model. *J. Dairy Sci.* 99, 427–442. doi:10.3168/jds.2015-9768

Wiggans, G. R., Cole, J. B., Hubbard, S. M., and Sonstegard, T. S. (2017). Genomic selection in dairy cattle: The USDA experience. *Annu. Rev. Anim. Biosci.* 5, 309–327. doi:10.1146/annurev-animal-021815-111422

Xiong, R., Yin, T., Gao, J. L., and Yuan, Y. F. (2020). HOXD9 activates the TGF-β/smad signaling pathway to promote gastric cancer. *Onco Targets Ther.* 13, 2163–2172. doi:10.2147/OTT.S234829

Xu, F., Shangguan, X., Pan, J., Yue, Z., Shen, K., Ji, Y., et al. (2021). HOXD13 suppresses prostate cancer metastasis and BMP4-induced epithelial-mesenchymal transition by inhibiting SMAD1. *Int. J. Cancer* 148, 3060–3070. doi:10.1002/ijc.33494

Xu, L., Shi, L., Liu, L., Liang, R., Li, Q., Li, J., et al. (2019). Analysis of liver proteome and identification of critical proteins affecting milk fat, protein, and lactose metabolism in dariy cattle with iTRAQ. *Proteomics* 19, e1800387. doi:10.1002/pmic.201800387

Ye, W., Xu, L., Li, Y., Liu, L., Ma, Z., Sun, D., et al. (2022). Single nucleotide polymorphisms of ALDH18A1 and MAT2A genes and their genetic associations with milk production traits of Chinese Holstein cows. *Genes (Basel)* 13, 1437. doi:10.3390/genes13081437

Yin, J., and Guo, Y. (2021). HOXD13 promotes the malignant progression of colon cancer by upregulating PTPRN2. *Cancer Med.* 10, 5524–5533. doi:10.1002/cam4.4078

Zhang, H., Liu, A., Wang, Y., Luo, H., Yan, X., Guo, X., et al. (2021a). Genetic parameters and genome-wide association studies of eight longevity traits representing either full or partial lifespan in Chinese holsteins. *Front. Genet.* 12, 634986. doi:10.3389/fgene.2021.634986

Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., et al. (2015). Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3 (Bethesda)* 5, 615–627. doi:10.1534/g3.114.016261

Zhang, Z., Fang, X., Xie, G., and Zhu, J. (2021b). Erratum: GATA3 is downregulated in HCC and accelerates HCC aggressiveness by transcriptionally inhibiting slug expression. *Oncol. Lett.* 21, 575. doi:10.3892/ol.2021.12836

Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., et al. (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS One* 9, e93017. doi:10.1371/journal.pone.0093017

Zhu, H., Dai, W., Li, J., Xiang, L., Wu, X., Tang, W., et al. (2019). HOXD9 promotes the growth, invasion and metastasis of gastric cancer cells by transcriptional activation of RUFY3. *J. Exp. Clin. Cancer Res.* 38, 412. doi:10.1186/s13046-019-1399-1

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fgene.2022.1092242

Check for updates

OPEN ACCESS

EDITED BY
Ibrar Muhammad Khan,
Fuyang Normal University, China

REVIEWED BY
Muhammad Zahoor Khan,
University of Agriculture, Dera Ismail Khan, Pakistan
Zheng Chen,
Jiangxi Agricultural University, China
Shanshan Yang,
Hebei Normal University of Science and Technology, China

*CORRESPONDENCE
Michael N. Romanov,
m.romanov@kent.ac.uk

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

RECEIVED 07 November 2022
ACCEPTED 27 December 2022
PUBLISHED 11 January 2023

CITATION
Kochish II, Titov VY, Nikonov IN, Brazhnik EA, Vorobyov NI, Korenyuga MV, Myasnikova OV, Dolgorukova AM, Griffin DK and Romanov MN (2023), Unraveling signatures of chicken genetic diversity and divergent selection in breed-specific patterns of early myogenesis, nitric oxide metabolism and post-hatch growth. *Front. Genet.* 13:1092242. doi: 10.3389/fgene.2022.1092242

COPYRIGHT
© 2023 Kochish, Titov, Nikonov, Brazhnik, Vorobyov, Korenyuga, Myasnikova, Dolgorukova, Griffin and Romanov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Unraveling signatures of chicken genetic diversity and divergent selection in breed-specific patterns of early myogenesis, nitric oxide metabolism and post-hatch growth

Ivan I. Kochish[1], Vladimir Yu. Titov[1,2], Ilya N. Nikonov[1], Evgeni A. Brazhnik[3], Nikolai I. Vorobyov[4], Maxim V. Korenyuga[1], Olga V. Myasnikova[1], Anna M. Dolgorukova[2], Darren K. Griffin[5] and Michael N. Romanov[1,5]*

[1]K. I. Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology, Moscow, Russia, [2]Federal Scientific Center "All-Russian Poultry Research and Technological Institute" of the Russian Academy of Sciences, Sergiev Posad, Moscow Oblast, Russia, [3]BIOTROF+ Ltd., Pushkin, St. Petersburg, Russia, [4]All-Russia Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia, [5]School of Biosciences, University of Kent, Canterbury, United Kingdom

Introduction: Due to long-term domestication, breeding and divergent selection, a vast genetic diversity in poultry currently exists, with various breeds being characterized by unique phenotypic and genetic features. Assuming that differences between chicken breeds divergently selected for economically and culturally important traits manifest as early as possible in development and growth stages, we aimed to explore breed-specific patterns and interrelations of embryo myogenesis, nitric oxide (NO) metabolism and post-hatch growth rate (GR).

Methods: These characteristics were explored in eight breeds of different utility types (meat-type, dual purpose, egg-type, game, and fancy) by incubating 70 fertile eggs per breed. To screen the differential expression of seven key myogenesis associated genes (*MSTN*, *GHR*, *MEF2C*, *MYOD1*, *MYOG*, *MYH1*, and *MYF5*), quantitative real-time PCR was used.

Results: We found that myogenesis associated genes expressed in the breast and thigh muscles in a coordinated manner showing breed specificity as a genetic diversity signature among the breeds studied. Notably, coordinated ("accord") expression patterns of *MSTN*, *GHR*, and *MEFC2* were observed both in the breast and thigh muscles. Also, associated expression vectors were identified for *MYOG* and *MYOD1* in the breast muscles and for *MYOG* and *MYF5* genes in the thigh muscles. Indices of NO oxidation and post-hatch growth were generally concordant with utility types of breeds, with meat-types breeds demonstrating higher NO oxidation levels and greater GR values as compared to egg-type, dual purpose, game and fancy breeds.

Discussion: The results of this study suggest that differences in early myogenesis, NO metabolism and post-hatch growth are breed-specific; they appropriately reflect genetic diversity and accurately capture the evolutionary history of divergently selected chicken breeds.

KEYWORDS

chicken, genetic diversity, divergent selection, breeds, early myogenesis, differential gene expression, nitric oxide oxidation, post-hatch growth

1 Introduction

As a relatively inexpensive source of quality animal protein in the form of meat and eggs, the rearing and use of poultry is a very important livestock production sector (Bogolyubsky, 1991; Romanov et al., 2009). Over several millennia of domestication, breeding and selection, various poultry breeds and varieties have been created by humans that are adapted to certain conditions of keeping and exploited economically for meat, eggs and other purposes (e.g., cock fighting and aesthetic needs). These breeds present a variety of phenotypic traits and, accordingly, can be classified based on their origin, phenotypes, selection targets and utility purpose (Abdelmanova et al., 2021; Larkina et al., 2021; Romanov et al., 2021). According to the traditional classification model (Bogolyubsky, 1991), the main classes of chicken breeds are meat, egg, dual purpose (i.e., meat-egg and egg-meat), game, and fancy (or ornamental). Moiseyeva et al. (2003) postulated an evolutionary model of breed formation with four main branches: egg (Mediterranean), meat (Asian), game, and Bantam ones. Larkina et al. (2021) proposed a phenotypic clustering model, supplementing the evolutionary model with two more breed types, i.e., dual purpose and fancy breeds (see breed examples in Table 1; Supplementary Material S2). Assessment of genetic diversity in various breeds is an important element in developing new strategies and applications for poultry breeding and production, as well as germplasm preservation (Romanov and Weigend, 1999; 2001; Huang et al., 2016; Romanov et al., 2017; Romanov et al., 2021; Bernini et al., 2021; Dementieva et al., 2021).

The manifestation of differences between divergently selected and genetically diverse poultry types and breeds can be expected at the earliest stages of embryonic and postnatal development. First of all, this can be traced by the breed-specific features of early myogenesis in embryos and postnatal growth in chicks as was shown by Kanakachari et al. (2022). However, that study included only a broiler line and a native Indian breed. Therefore, it would be reasonable and purposeful to establish the respective genetic diversity signatures by comparing differential gene expression (DGE) among genes responsible for myogenesis in muscle tissues in a broader sample of various chicken breeds divergently selected for meat and egg performance and other phenotypic traits.

Cazzato et al. (2014) studied the earliest stages of embryogenesis and DGE for five key genes controlling the course of skeletal muscle development, such as myosin (*MYH1*; e.g., Thompson et al., 2021) and related ones. As shown by Cazzato et al. (2014), there are effects of nitric oxide synthase inhibitor (NOSI) and nitric oxide donor (NOD) compounds on DGE of myogenesis associated genes. To date, more evidence has been accumulated regarding the role of nitric oxide (NO) in embryogenesis and, in particular, myogenesis (e.g., Cazzato et al., 2014; Titov et al., 2018; Titov et al., 2020b; Titov et al., 2021; Dolgorukova et al., 2020). NO is believed to mediate myocyte proliferation (Ulibarri et al., 1999; Stamler and Meissner, 2001; Long et al., 2006; Li et al., 2016; Tirone et al., 2016), muscle fiber formation (Stamler and Meissner, 2001; Long et al., 2006), and satellite cell proliferation (Anderson, 2000). In conformity with current concepts, the physiological effect of NO manifests itself through the nitrosation of certain protein structures/enzymes: guanylate cyclase (Stamler et al., 1992; Severina et al., 2003), caspases (Dimmeler et al., 1997; Rossig et al., 1999; Kim et al., 2000), as well as molecular cellular factors that determine transcriptional regulation and DGE (Zhou and Brüne, 2005; Vasudevan et al., 2016; Socco et al., 2017).

To evaluate the role of NO in a particular physiological process, a monitoring technique for its synthesis and metabolism to the final product, nitrate, is needed. Synthesized NO is included in NOD compounds (Titov et al., 2020a): S-nitrosothiols (RSNO), dinitrosyl-iron complexes (DNIC), and high molecular weight nitro derivatives (RNO_2). These compounds play the role of physiological depots of NO, prolonging its physiological lifetime (Severina et al., 2003; Vanin et al., 2017; Vanin et al., 2017). Their concentration in cells can reach tens of μM (Hickok et al., 2011; Titov et al., 2016). Therefore, to determine the NO role in a specific process (e.g., embryogenesis), it is necessary to monitor changes in the content of deposited NO and its metabolic products during this process. It is not straightforward to precisely detect the content of all NO metabolites in living tissues, e.g., methods for determining DNIC and RSNO do not have high accuracy and specificity (Tarpey et al., 2004; Titov, 2011; Vanin et al., 2017; Vanin et al., 2017). To address this problem, conclusions about the effect of NO on a particular physiological process can be inferred based on the effects of NOSI, NOD compounds, and arginine, which is a source of NO (Stamler et al., 1992; Anderson, 2000; Long et al., 2006; Cazzato et al., 2014).

Previously, we developed an enzymatic sensory method that is based on reversible inhibition of catalase by all nitroso compounds and enables detecting the concentration of RSNO, DNIC, nitrite, and nitrosamines with an accuracy of 50 nM (Titov, 2011; Titov et al., 2016). Using this sensor, we confirmed an assumption that DNIC are the main NOD in most tissues (Titov et al., 2016; Titov et al., 2018). It was shown that the embryogenesis of birds, like in other animals, is associated with intense production of NO that either accumulates in the embryo as part of NOD compounds or is oxidized to nitrate. NO oxidation proceeds throughout the entire embryonic period. Within the same species, the intensity of NO synthesis is approximately the same but there are differences in the degree of NO oxidation to nitrate. The latter indicator, according to our previous observations (Titov et al., 2018; Titov et al., 2021), can be many times higher in meat-type chickens than in egg-type breeds. Post hatch, the concentration of nitro compounds and nitroso compounds in the chick tissues declines sharply as compared with the embryo tissues and levels off in various breeds, lines, and crosses (Titov et al., 2018). Analysis of the content of nitro compounds and nitroso compounds in various embryo tissues showed that nitrate mainly accumulates in muscle tissue. It does not accumulate in the liver and intestines to any great degree (Titov et al., 2018), and apparently, NO oxidation mainly occurs in the muscle tissue.

The present study aimed to explore signatures of chicken genetic diversity and divergent selection by examining breed-specific patterns of early myogenesis (assessed by DGE of myogenesis associated genes) and post-hatch growth in various breeds. Therewith, one of the objectives was also to investigate mechanisms of the relationship between the utility type of chicken breeds and intensity of NO oxidation in embryos (i.e., in their muscle tissues) among various breeds.

2 Materials and methods

2.1 Chicken breeds and sampling

In this investigation, eight chicken breeds and crosses were used (Table 1), which were kept in grower cages and fed following

TABLE 1 Characterization of the studied chicken breeds.

Breed	Code	Type of divergent selection			Origin	Description
		TCM[a]	EM[b]	PCM[c]		
Broiler	BR	Meat	Meat	Meat	Russia	4-way BR cross Smena 8 developed in 2011 at the Breeding Genetic Center "Smena"—Branch of the Federal Scientific Center "All-Russian Poultry Research and Technological Institute" of the Russian Academy of Sciences
White Cornish	WC	Meat	Meat	Meat	Russia/England	B56, male grandparent stock, of BR cross Smena 8. The initial breed was developed from English local game chickens, Asil, White Malay, Indian Game, and Cochin
Plymouth Rock White	PRW	Dual purpose (meat-egg type)	Meat	Dual purpose (meat-egg type)	Russia/ United States	B79, female grandparent stock, of BR cross Smena 8. The initial breed was developed from Java Black, Brahma, Cochin White and Buff, Dominique, and White-faced Black Spanish
Yurlov Crower	YC	Dual purpose (meat-egg type)	Meat	Dual purpose (meat-egg type)	Russia	Derived in 19th century from crossing local and game chickens, Brahma, Cochin, and Langshan. Selected for long crowing
Brahma Buff	BB	Fancy	Meat	Dual purpose (egg-meat type)	USA/India	Derived in early 20th century from crossing Cochin and Gray Chittagong (of Malay type)
Orloff Mille Fleur	OMF	Fancy/Game	Game	Dual purpose (meat-egg type)	Russia	Derived in late 18th century from crossing local chickens, Gilan, and Old English Game
Layer	LR	Egg	Egg	Egg	Netherlands	Commercial 4-way layer cross Hisex White
Uzbek Game (Kulangi)	UG	Game	Game	Game	Uzbekistan	An old cock fighting breed derived from local Uzbek game chickens

Chicken breed types according to:
[a]TCM, traditional classification model (Bogolyubsky, 1991).
[b]EM, evolutionary model (Moiseyeva et al., 2003).
[c]PCM, phenotypic clustering model (Larkina et al., 2021).

recommendations as prescribed by the Federal Scientific Center "All-Russian Poultry Research and Technological Institute" of the Russian Academy of Sciences (Imangulov et al., 2013). Seventy fertile hatching eggs per breed were used for incubation, while the proper embryo samples were analyzed at embryonic age of 7 (E7) and 14 (E14) days. The content of NO metabolites was determined in embryos at E7, and the DGE level of myogenesis associated genes in the tissues of the breast and thigh muscles was assessed at E14. Incubators Stimul Ink-1000 (OOO Stimul Group, Russia) were used for incubation. Temperature was 37.6°C during the incubation period and 37.2°C at hatching. To obtain homogenates from the whole E7 chick embryos (four per breed), the egg contents were used after removing the eggshell. The contents were processed in a glass homogenizer for 8 min at 40 fpm and 6°C. A tissue grinder was used to obtain breast and thigh muscle tissue homogenates at E14 followed up by total RNA isolation using the RNeasy Midi Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions.

2.2 DGE assessment

Relative DGE levels of myogenesis associated genes were examined in the tissues of the breast muscles (Table 2) and thigh muscles (Table 3) in at least five E14 chick embryos per breed (with three technical replicates per sample). Using quantitative real-time PCR and sets of gene-specific primer pairs described elsewhere (e.g., Cazzato et al., 2014; Supplementary Table S1), we analyzed DGE for the following seven genes: *MSTN*, myostatin; *GHR*, growth hormone receptor; *MEF2C*, myocyte enhancer factor 2c; *MYOD1*, myogenic differentiation 1; *MYOG*, myogenin; *MYH1*; and *MYF5*, myogenesis factor 5. For internal DGE control, the TATA-binding protein (*TBP*) gene was used.

Based on the results of DGE assessment (Tables 2, 3), a Type I dataset was formed, which included raw values of fold change (FC) as a derivative of C_T (Livak and Schmittgen, 2001; Schmittgen and Livak 2008). If there was upregulated DGE of a gene relative to the internal control gene, i.e., when $\Delta C_T < 0$, FC value was calculated using the following formula: $FC = 2^{-\Delta C_T}$. In the case of downregulated DGE of a gene relative to the internal control gene, i.e., when $\Delta C_T > 0$, FC was determined using the following formula: $FC = \frac{-1}{2^{-\Delta C_T}}$ (Schmittgen and Livak 2008). Subsequently, four datasets were used for DGE assessment, including one raw FC dataset (I) and three transformed datasets (II to IV). Herewith, appropriate normalizing algorithms were applied for calculating values of shifted DGE levels relative to each other, so that the numbers resulted from mathematical transformation (normalization) were more adequate and convenient for further mathematical processing and analyses (see Supplementary Information SI1 for further details).

2.3 Estimation of embryonic NO oxidation rate

The content of NO metabolites in the E7 embryo samples was tested no later than 30 min after sampling. We used the enzymatic

TABLE 2 Relative DGE levels defined by raw FC values (type I data) in the breast muscle tissues of E14 chick embryos as estimated in the studied breeds.

Breeds	Genes[a]						
	MSTN	*GHR*	*MEF2C*	*MYOD1*	*MYOG*	*MYH1*	*MYF5*
Broiler	11.55	6.63	6.59	11.31	7.46	−41,760.00	−7.57
White Cornish	4.89	5.62	2.91	2.19	−4.32	−16.22	−685.02
Plymouth Rock White	6.59	4.35	4.00	2.87	78.25	−24.42	−4.76
Yurlov Crower	121.9	69.1	302.3	−7.11	2.04	1.07	−5.90
Brahma Buff	41.07	31.78	219.8	−25.46	−1.95	−1.73	−8.57
Orloff Mille Fleur	2.41	3.32	2.33	16.11	5.58	−16,270.00	−37.53
Layer	4.72	4.79	4.14	4.59	1.03	−29.45	−66.26
Uzbek Game	1.18	2.51	1.45	−81.01	−106.9	−11,990.00	−4.47

[a]*MSTN*, myostatin; *GHR*, growth hormone receptor; *MEF2C*, myocyte enhancer factor 2C; *MYOD1*, myogenic differentiation 1; *MYH1*, myosin heavy chain 1; *MYOG*, myogenin; *MYF5*, myogenic factor 5. Internal control gene used: *TBP*, TATA-box binding protein.

TABLE 3 Relative DGE levels defined by raw FC values (type I data) in the thigh muscle tissues of E14 chick embryos as estimated in the studied breeds.

Breeds	Genes[a]						
	MSTN	*GHR*	*MEF2C*	*MYOD1*	*MYOG*	*MYH1*	*MYF5*
Broiler	3.86	3.07	2.36	18.77	6.73	−10,020.00	−6.45
White Cornish	4.03	3.05	−1.69	−12.13	−4640.29	−335.46	−25,531.63
Plymouth Rock White	4.50	2.95	−1.02	13.18	1.39	−115.36	−33.36
Yurlov Crower	46.53	26.10	494.56	28.44	1.39	2.30	195.36
Brahma Buff	8.86	3.72	63.39	8.78	−2.70	1.37	38.02
Orloff Mille Fleur	−1.28	1.62	1.78	6.06	3.63	−8,481.00	−18.90
Layer	1.25	1.31	2.46	1.08	−78.25	6.23	−87.43
Uzbek Game	3.25	4.92	−4.79	−13.93	−118.60	−17,560.00	−2.43

[a]*MSTN*, myostatin; *GHR*, growth hormone receptor; *MEF2C*, myocyte enhancer factor 2C; *MYOD1*, myogenic differentiation 1; *MYH1*, myosin heavy chain 1; *MYOG*, myogenin; *MYF5*, myogenic factor 5. Internal control gene used: *TBP*, TATA-box binding protein.

sensor we previously developed (Titov, 2011; Titov et al., 2016). Its detecting sensitivity is based on property of nitrite, nitrosamines (RNNO), RSNO, DNIC, and RNO_2 to inhibit catalase in the presence of halide ions and on their loss of this property under the influence of factors different for each group of compounds. The nitrate content was estimated after reduction with vanadium trichloride to nitrite followed by quantitative determination (Titov, 2011). The enzymatic sensor is designed using a highly sensitive calorimeter Dithermanal (Vaskut-EMG, Hungary). Since the catalase process is highly exothermic (47.2 kcal/1 mol of released oxygen), its kinetics can be monitored by the kinetics of heat production accompanying this process (Titov, 2011; Titov et al., 2016). This method enables estimating the content of NO derivatives without preliminary sample preparation, since there is no need to remove colored impurities and turbidity in samples. The sensor sensitivity is up to 50 nM (Titov, 2011; Titov et al., 2016). The classical Griess reaction method was also used to determine nitrite (Tarpey et al., 2004). DNIC containing two glutathione (GSH) molecules was used as NOD according to the technique we previously described (Titov, 2011; Titov et al., 2016). Solutions prepared in sterile saline were administered *in ovo* before incubation using injection into the air cell of the egg.

2.4 Analysis of embryonic development and postnatal growth

For a comparative assessment of the features of embryonic development and postnatal growth in chicks of various breeds, the following indicators were measured: mean weight of fertile eggs prior to incubation, body weight (BW) of chicks at three ages (1, 14, and 28 days) and the degree of NO oxidation to nitrate in the homogenates of E7 embryos. To expand the representative set of various breed types, the following breeds/crosses were also added to the eight initial breeds: two BR crosses, Cobb 500 (BRC) and Ross 308 (BRR); one game breed, Malay Game; and two dual purpose breeds, Andalusian Blue and Blue Meat-Egg Type (BMET) that was selected from the Andalusian Blue breed. A total of 13 chicken breeds were used within this research phase (Supplementary Table S2). Postnatal growth rate (GR) due to the growth of skeleton and muscles, primarily the breast and thighs, was estimated by the degree of BW gain over the first 2 and 4 weeks of life. Accordingly, GR was calculated for 2 weeks (GR2wk) and 4 weeks (GR4wk) by dividing the respective values of BW at 2 and 4 weeks by BW at day old. Further, we also tested relationship between DGE levels of myogenesis associated genes assessed in E14 embryos and GR2wk/GR4wk values.

2.5 Principal component analysis, clustering, and statistical processing

Principal component analysis (PCA) and PCA plotting were performed in RStudio (version 1.1.453; RStudio Team, 2016) using the ggplot2 library (version 3.3.5; Wickham, 2009; Pedersen, 2021; Wickham et al., 2021). In addition, PCA plots were built using the web toolbox ClustVis (Metsalu and Vilo, 2015) designed for visualizing clustering of multivariate data. PCA plots were originally obtained by applying the unit variance scaling to rows of the original Type I raw data matrix (with preserving their signs). To calculate principal components, multilevel singular value decomposition with imputation was used. Heat maps and their accompanying clustering trees were built using ClustVis and Euclidean distances for both rows and columns of the matrix (with the average option selected for the linkage method). Additionally, PCA and hierarchical clustering procedures were employed using the Phantasus web application (Zenkova et al., 2018).

Hierarchical clustering was also performed using the pvclust package in R (Suzuki and Shimodaira, 2006). For clustering, the Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA) was applied using the Euclidean distance measure. Bootstrapping with 10,000 iterations was implemented for validation. Using the fviz_nbclust() function from the factoextra package (Kassambara and Mundt, 2017), optimal number of clusters was chosen using the elbow method (Zhao et al., 2008). To test significance of the UPGMA-based hierarchical clustering output, agglomerative coefficient that measures magnitude of the clustering structure found (values close to 1 suggest a strong clustering structure) was calculated using the "agnes" function from the "cluster" package (version 2.1.2; Maechler et al., 2021). Trees of phylogenetic relationships between breeds were constructed using the Neighbor Joining method (Saitou and Nei, 1987) using the online T-REX tool (Boc et al., 2012). Two options were used to build trees: radial topology 1) with proportional lengths of edges and 2) without it.

Manipulations of normal mathematical processing of primary data were carried out using MS Excel. In addition, BioStat software package and RStudio (version 1.1.453; RStudio Team, 2016) were also used for statistical analyses. To assess distribution normality of quantitative traits, the Shapiro–Wilk test was applied using the base function shapiro.test() for R. Since the data did not have a normal distribution, correlation analysis was performed using the Spearman's rank-order correlation test and the base function cor() for R. Data visualization was performed using the corrplot package (version 0.90) for R (Wei and Simko, 2021).

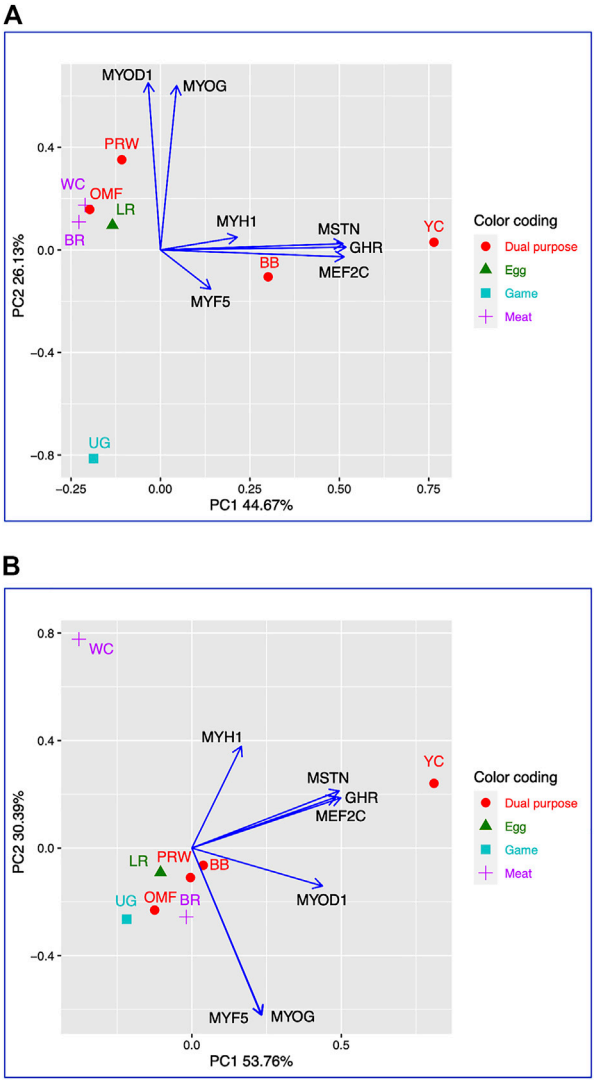


FIGURE 1
PCA plots generated using the ggplot2 library for the DGE Type I data for the seven genes in the breast **(A)** and thigh **(B)** muscles in the eight breeds studied. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 44.7% and 26.1% **(A)**, and 53.8% and 30.4% **(B)** of the total variance, respectively. *N* = 8 data points (breeds).

3 Results

3.1 Analysis of DGE patterns of myogenesis associated genes in different chicken breeds

3.1.1 Relative DGE data

As follows from the Type I data for the relative DGE in muscle tissues of E14 embryos (Tables 2, 3), essential (i.e., two-fold and higher) upregulation and downregulation of the seven myogenesis associated genes studied were observed in various breeds. For example, WC had markedly lower FC values for the *MYF5* gene in both breast and thigh muscles. Considering raw Type I data (Tables 2, 3), one cannot directly identify any apparent pattern in the DGE levels among the breeds studied. At first glance, each breed was characterized by its own combination of up and downregulation of certain genes.

3.1.2 Cluster analysis of DGE patterns

Using for analysis the available raw Type I data matrices for DGE obtained for the seven genes in the breast (Table 2) and thigh (Table 3) muscles in the eight breeds, the clustering structure of the differentially expressed genes (DEG) was analyzed in more detail (Figure 1). Thereby, PCA plots built in R environment using the ggplot2 library demonstrated the DEG clustering patterns for the breast (Figure 1A) and thigh (Figure 1B) muscles, suggesting, in a first approximation, the effects and interactions of DEG in vector form. Gene vectors in the PCA plots (Figure 1) suggested that three genes (*MSTN*, *GHR*, and *MEFC2*) were expressed as if by one "accord" (complex), i.e., interconnected and in one direction, both in the breast

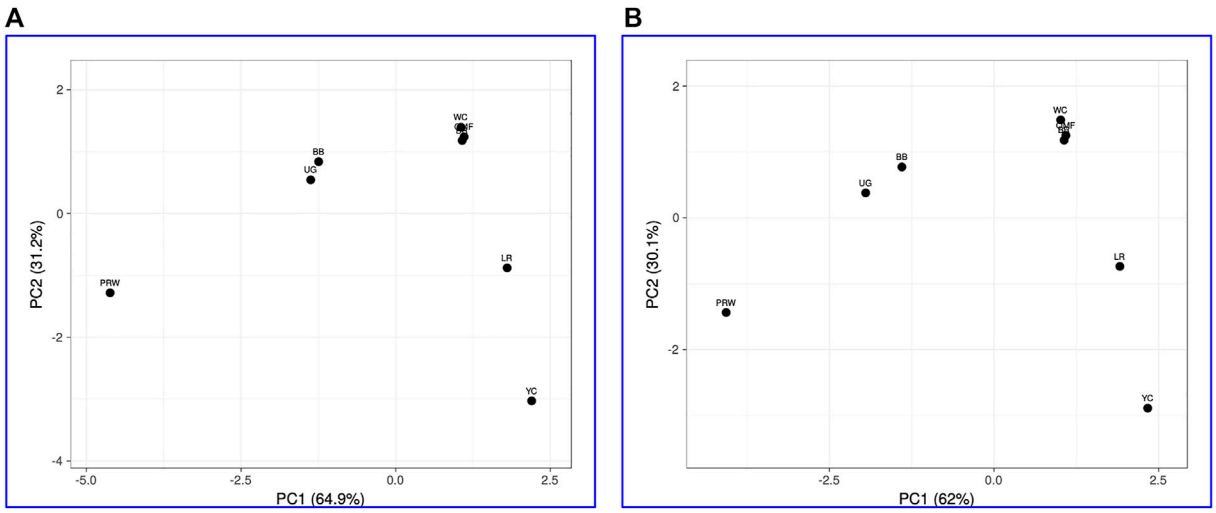


FIGURE 2
PCA plots generated using the ClustVis tool (Metsalu and Vilo, 2015) and the Type IIIa **(A)** and IIIb **(B)** datasets as inferred for the eight studied breeds and seven tested myogenesis associated genes expressed in the breast muscles. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 64.9% and 31.2% **(A)**, and 62.0% and 30.1% **(B)** of the total variance, respectively. *N* = 8 data points (breeds).

(Figure 1A) and thigh (Figure 1B) muscles. In other words, it seems that the *GHR* gene was one of the key ones and was associated with other genes forming a complex. There were also associated and unidirectional vectors for the *MYOG* and *MYOD1* genes expressed in the breast muscles (Figure 1A) and the *MYOG* and *MYF5* genes in the thigh muscles (Figure 1B). The identified "accord" DGE patterns during early muscle development were mainly confirmed when using other variants of PCA analysis and hierarchical clustering (see Supplementary Information SI2 for further details).

3.1.3 Cluster analysis of the studied breeds based on DGE data

When analyzing distribution of the eight breeds studied based on DGE data in the breast (Table 2) and the thigh (Table 3) muscles, meaningful clustering patterns were obtained, especially after transformation of primary raw data as outlined below (see also Supplementary Information SI3 for further details). In particular, when using transformed DGE matrices to build PCA plots using the ClustVis web service, similar clustering patterns were obtained based on both Type IIIa (Figure 2A) and Type IIIb (Figure 2B) datasets for the breast muscles. Unlike the raw data I based clustering pattern (Figure 1A), the egg-type LR breed was significantly moved out of the crowded core of breeds in these PCA plots (Figures 2A, B). In general, the PCA results inferred using ClustVis repeated the Neighbor Joining analysis outputs (see Supplementary Information SI3; Supplementary Figure SI3-4).

Analysis of the transformed data (i.e., two matrices for the Type IIIa and Type IIIb datasets) was further performed using the Phantasus web service (Zenkova et al., 2018) and the standard PCA algorithm (Figure 3). Additionally, we tested slightly changed the Type IIIa and IIIb datasets through obtaining similarity matrices based on the Euclidean metric (Supplementary Figure S1). The obtained PCA plots for the breast muscles (Figures 3A, B; Supplementary Figures S1A, B) basically repeated the patterns of PCA clustering (Figure 2) obtained using ClustVis (Metsalu and Vilo, 2015). For example, LR again turned out to be distanced from the crowded core of other breeds with a shift to the right plot side. Moreover, the crowded core itself in Supplementary Figures S1A, B was represented by two breeds, WC and OMF, and the BR breed moved away from it to the left. It seems that these PCA plots more plausibly described the DGE results obtained for early myogenesis associated genes in the breast muscles in various chicken breeds studied. In the case of the thigh muscles (Figures 3C, D; Supplementary Figures S1C, D), we had crowded cores of six different breeds with two other breeds being remote from them. At the same time, the apartness of the meat-type (WC) and egg-type (LR) breeds in Figures 3C, D looked most plausible. Note also that there was practically no difference in PCA patterns between the two data types, IIIa (Figures 3A, B; Supplementary Figures S1A, B) and IIIb (Figures 3C, D; Supplementary Figures S1C, D). Overall, the obtained PCA plots (Figure 3; Supplementary Material S1; cf. also the increased sums of the proportions of PC1 and PC2 for the breast and thigh muscles as compared to those for PCA using raw Type I data in Supplementary Information S4) indicated a slightly better resolution of the PCA method in comparison with the Neighbor Joining analysis results (see Supplementary Information SI3; Supplementary Figure SI3-4). Next, another Phantasus option was used, i.e., hierarchical clustering based on the Euclidean metric (Figure 4). In this case, two data types, IIIa (Figure 4A) and IIIb (Figure 4B), were also compared and the resulting trees were identical to each other. This means that both approaches to the transformation of raw data did not contradict each other. According to the character of DGE in the breast muscles, the examined breeds were divided into four main clusters: 1) two meat-type breeds (WC and BR) and one dual purpose breed (OMF); 2) two dual purpose breeds (YC and BB) and one related (by descent) game breed UG; 3) the egg-type LR breed; and 4) the dual purpose PRW breed (female grandparent stock of the BR cross). To improve the sensitivity of hierarchical clustering, both datasets, IIIa and IIIb, were also preliminarily subjected to the procedure for constructing similarity matrices using the Euclidean distance

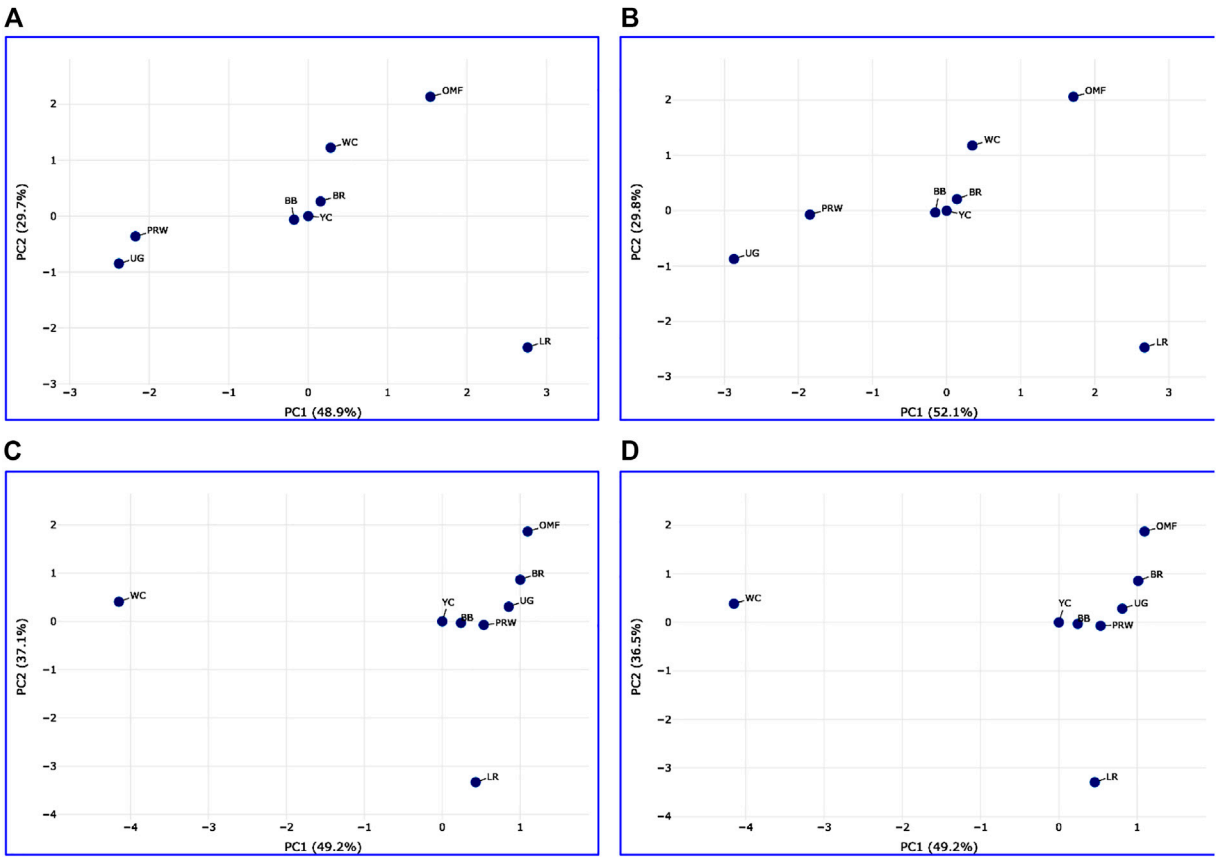


FIGURE 3
PCA plots generated using the Phantasus tool (Zenkova et al., 2018) and the data Type IIIa **(A,C)** and IIIb **(B,D)** as inferred for the eight studied breeds and seven tested myogenesis associated genes expressed in the breast **(A,B)** and thigh **(C,D)** muscles. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain respective percentage values of the total variance. *N* = 8 data points (breeds).

metrics (Figures 5A, B) and Pearson's correlation coefficient (Figures 5C, D). The generated clustering patterns for Type IIIa data (Figures 5A, C) were respectively identical to those for Type IIIb counterparts (Figures 5B, D).

For DEG in the breast muscles, five main clusters were identified (Figures 5A, B): 1) the one coinciding with the first cluster in Figure 4 and splitting into two subclusters BR and WC + OMF; 2) the one combining BB and UG; 3) LR; 4) YC; and 5) PRW. For the thigh muscles, we had a completely different pattern of hierarchical clustering: 1) one large cluster with two dual purpose breeds, YC and BB, and two BR grandparent stocks, WC and PRW; 2) the meat-type BR breed; 3) the game UG breed and OMF considered as a "semi-game" breed; and 4) the egg-type LR breed. The clustering trees obtained for the breast and thigh muscles (Figure 5) were even more demonstrative than the PCA patterns (Figures 2, 3; Supplementary Material S1). Moreover, these trees showed phylogenetic relationships much better than the Neighbor Joining trees (see Supplementary Information SI3; Supplementary Figure SI3-4).

On the whole, the results of analyzes using PCA and hierarchical clustering suggest the peculiar nature of DGE patterns in both breast and thigh muscles in embryos of two breeds, imported egg-type LR and domestic dual purpose YC, which distinguished these breeds from the rest. At the same time, the proximity of YC to another dual purpose breed, BB, was also noted. In the breast muscles, a distinct DGE pattern was observed in the meat-type BR breed, and it was close to the patterns in meat WC and dual purpose OMF. The dual purpose PRW breed (female grandparent of the BR cross) was also distinguished by the DGE peculiarity in this tissue. We can also suggest the similarity of DGE patterns in the breast muscles in game UG and dual purpose BB. In addition, in the thigh muscles, we also observed a peculiar DGE pattern in meat WC (male grandparent of the BR cross).

3.2 Analysis of embryonic development and postnatal growth in various chicken breeds

Indicators of the mean egg weight (EW), BW of chicks at three ages (1, 14, and 28 days), and the degree of NO oxidation to nitrate in the homogenates of E7 embryos for the 13 chicken breeds are presented in Supplementary Table S2. Significant interbreed variability in values of the studied traits was noted.

3.2.1 Early growth traits

Based on the growth data obtained (Supplementary Table S2), it was feasible to identify breeds with approximately similar GR. For instance, breeds such as BR cross Smena 8 (BRS), WC and PRW, or a

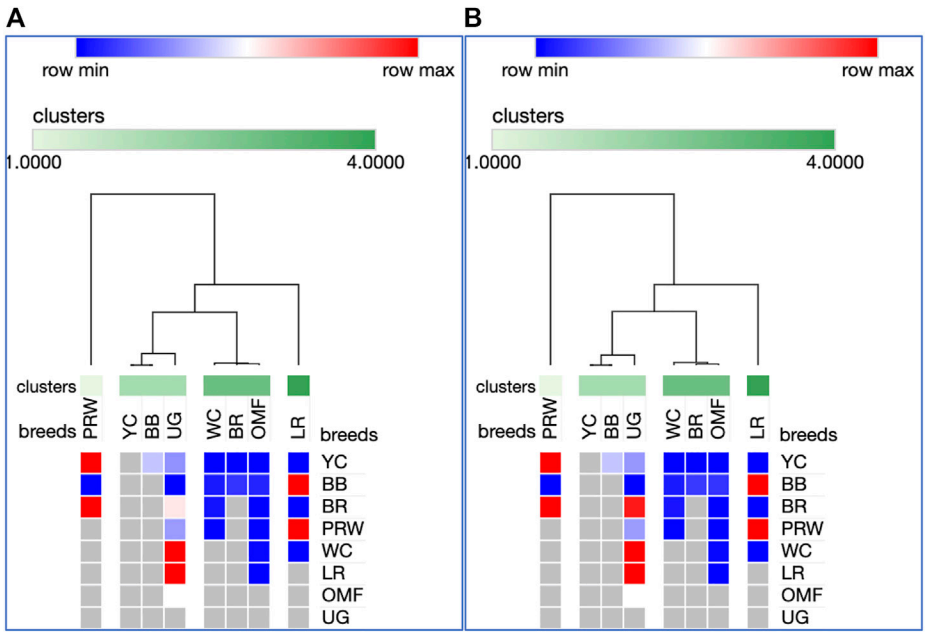


FIGURE 4
Heatmaps and hierarchical clustering trees based on Euclidean distance metric (with the average option selected for the linkage method) and using the Type IIIa **(A)** and IIIb **(B)** data as inferred for the eight studied breeds and seven tested myogenesis associated genes expressed in the breast muscles.

pair of dual purpose breeds, YC and BB had similar values of BW at 1-, 14- and 28-days of age, respectively. Overall, when assessing the GR values, the three BR crosses, BRS, BRC and BRR, as well as their grandparent stocks, WC and PRW, were expectedly similar. Their GR2wk values ranged from 5.70 to 7.87. The same indicators in game breeds, MG and UG, were 2.24 and 2.61, respectively, in dual purpose breeds they ranged from 2.01 to 2.81, and in LR it was the lowest (1.88). The GR4wk values were again maximum in BR breeds and their grandparent stocks (23.57–26.41), whereas it was 5.42 and 5.67 in game breeds, 4.59 to 6.33 in dual purpose breeds, and 5.25 in LR.

Further, we deduced and explored patterns of the embryonic and postnatal development in various chicken breeds using PCA and hierarchical clustering analyzes. Based on the indicators of EW (i.e., at the initial point of embryo development) and BW of chicks at three ages (i.e., at three temporal points of post-hatch development), plots in Figures 6A, B suggested the formation of two large clusters occupying respectively, the left and right parts of both graphs: one was a BR cluster (three crosses and two grandparent stocks), and the other consisted of the rest of the breeds. As demonstrated in the PCA plot (Figure 6A), the egg-type LR breed occupied a somewhat remote position at the bottom of the plot. Interestingly, a closely related pair of two dual purpose breeds, YC and BB, formed a separate cluster on the right side, as was also seen for DGE of embryonic myogenesis associated genes (see Figures 3, 4, 5C, D; Supplementary Information SI2; Supplementary Figures SI2-1C, D; Supplementary Information SI3; Supplementary Figures SI3-2, 3, 5C, D). A close pair was also made up by the two game breeds, MG and UG. In addition, in the upper right corner of the PCA plot (Figure 6A), two closely related dual purpose breeds, AB and BME, were located next to each other. In many respects, a similar pattern of breed clustering was observed on the corresponding heat maps (Figure 6A; Supplementary Material S2). Almost the same distribution patterns of the 13 breeds were obtained by adding GR indices to the set of analyzed growth traits (Supplementary Figure S3).

3.2.2 NO exchange in relation to early chick growth

When considering the data on the degree of NO oxidation to nitrate in homogenates of E7 embryos in the 13 chicken breeds (Supplementary Table S2, with the breeds being sorted in descending order by the level of NO oxidation), it can be seen that the highest NO oxidation values were inherent in three BR crosses (BRS, BRC, and BRR), meat-type WC (male grandparent stock) and two game breeds (MG and UG) and accounted for 96.9%–98.1%. High values of this indicator were also observed in two dual purpose breeds, BME and BB (61.8 and 74.1, respectively). The rest of the breeds had contrastingly low values of the NO oxidation degree (2.6% and below). Next, we tested the relationship between the level of embryonic NO metabolism and early postnatal growth of chickens. To do this, the clustering patterns of breeds were checked altogether for traits of NO oxidation at E7 and BW at day old (Supplementary Figure S4). The formed three clusters conformed to the same differences revealed by the level of NO oxidation (Supplementary Table S2), which was understandable since the day-old chicks did not differ much from each other in BW.

As the chicks grew, there were changes in interbreed differences. At 14 days of age, somewhat different clustering patterns were already observed (Supplementary Figures S5, S6). In particular, a pair of game breeds, MG and UG, and two dual purpose breeds, BME and BB, moved away from the BR cluster. The breeds seen in the right cluster in Supplementary Figure S4 formed a crowding pattern, with PRW detached from them and moved closer to the BR cluster. The same patterns of interbreed differences can be generally noted for the entire observation period, i.e., up to 28 days of age of the chicks (Figures 7A-C; Supplementary Material S7). When modifying the hierarchical

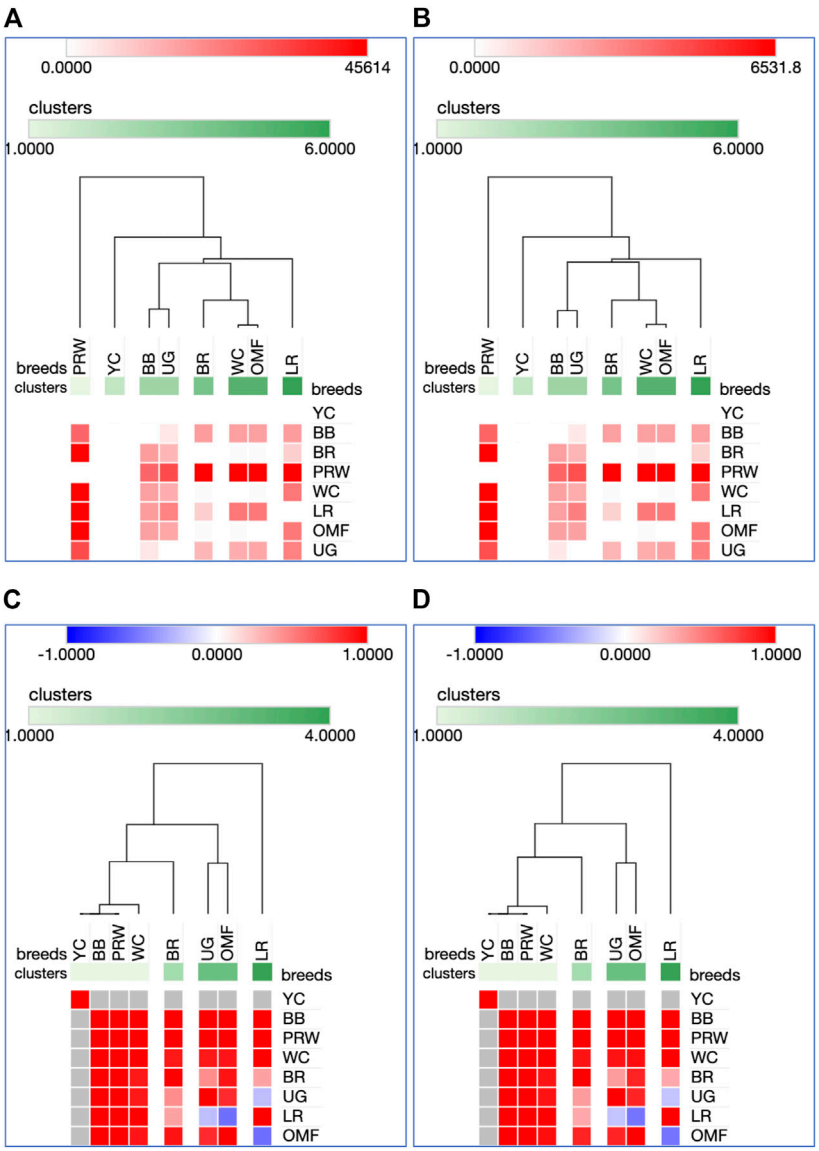


FIGURE 5
Heatmaps and hierarchical clustering trees based on Euclidean distance metric (with the average option selected for the linkage method) and using the Type IIIa **(A,C)** and IIIb **(B,D)** data as inferred for the eight studied breeds and seven tested myogenesis associated genes expressed in the breast muscles. For constructing the trees, precomputed similarity matrices were built using Euclidean distance **(A,B)** and Pearson correlation **(C,D)** as metrics.

clustering by applying the One minus Pearson's correlation metric (using the average linkage method; Figure 7C), the chicken breeds under consideration were divided into three clusters according to their utility purpose: 1) game (MG and UG); 2) meat-type (BRS, BRR, BRC, WC, PRW plus two dual purpose breeds, BB and BME); and 3) the remaining dual purpose breeds and the egg-type LR breed. Notably, when using this option of hierarchical clustering, we obtained a slightly different pattern of effects of the studied traits of early development and growth (Figure 7C): the indicators of the weight of fertile eggs and BW at 1, 14, and 28 days of life were isolated into a separate cluster detached from the E7 NO oxidation index (as also seen in the plot of Supplementary Figure S7).

The data used for analyzing the same five traits (including the E7 NO oxidation) correlated well with each other using the Spearman's correlation coefficient (Supplementary Table S3; Supplementary Figure S8). The E7 NO oxidation had a significant and positive correlation with the indicators of the BW of chickens, starting from hatching and up to 28 days of age. By adding two more parameters to this set of traits, GR2wk and GR4wk, more differentiated patterns of relationships between breeds were obtained (Supplementary Figures S9, S10). Remarkably, if we look at the relationships of all seven studied traits of embryonic NO metabolism and early growth of chicks (Supplementary Figures S9B, S11), we can see that GR2wk and GR4wk almost coincide with each other, suggesting their almost equal contribution to the observed clustering patterns of the 13 breeds. EW and BW of day-old chicks were close to each other that can be indicative of insignificant interbreed differences in these indicators of the earliest chicken development. Further, this cluster in Supplementary Figure S9B joined with the E7 NO oxidation index, which is also

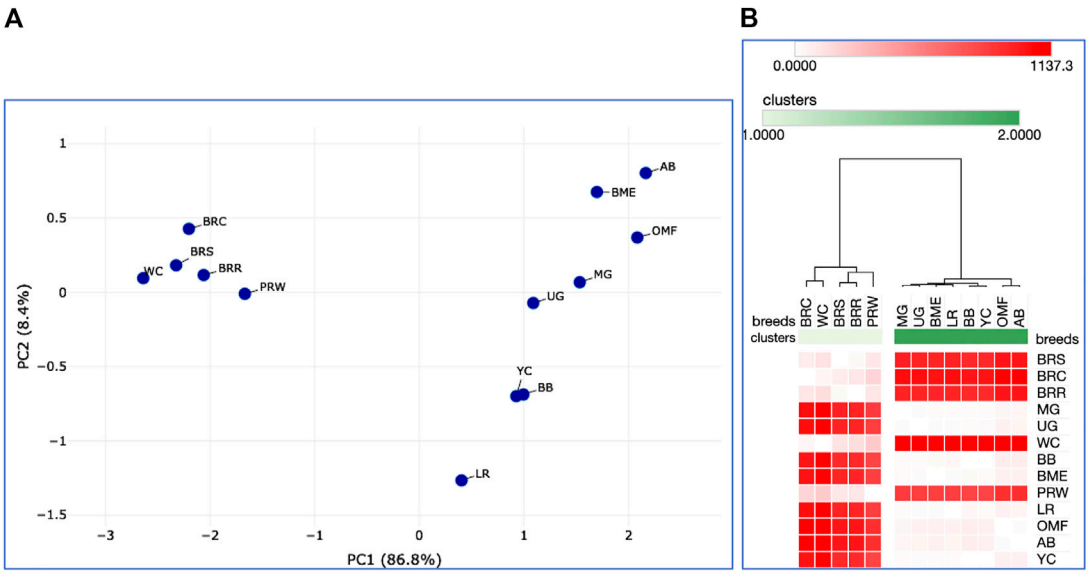


FIGURE 6
Analysis of the distribution of 13 breeds by early growth traits, including EW and BW of chicks at three ages, as generated in the Phantasus program (Zenkova et al., 2018). **(A)** PCA plot. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 86.8% and 8.4% of the total variance, respectively. *N* = 13 data points (breeds). **(B)** Heatmap and hierarchical clustering tree using Euclidean distance-based similarity matrix. For clustering, matrix values (for a precomputed distance matrix) were applied as metrics (with the average linkage method option selected).

understandable, since these three traits characterize embryonic development in the studied breeds. BW at 14 and especially at 28 days of age made a much stronger contribution to the pattern of interbreed differences, outlining specific trajectories in the further development and growth of birds of a particular breed.

3.2.3 Unified early development and growth model

We also tested a model embracing all various traits studied, i.e., DGE of myogenesis associated genes, NO metabolism in embryos, as well as seven indicators of early chick growth in the eight breeds. At first, using DGE indices for the seven genes, i.e., *MSTN*, *GHR*, *MEF2C*, *MYOD1*, *MYOG*, *MYH1*, and *MYF5*, in the tissues of the breast (Table 2) and thigh muscles (Table 3) in E14 chick embryos, the respective pairwise Spearman's correlation coefficients were calculated (Supplementary Tables S4, S5; Supplementary Figure S12). As a result of this analysis, significant pairwise correlation values were found between DGE indicators of some genes and between growth indicators. In particular, a significant correlation was confirmed between DGE levels for the three genes *GHR*, *MSTN*, and *MEF2C* in the breast muscles (Supplementary Table S4; Supplementary Figure S12A). This may reflect their key role in the embryonic myogenesis of the breast muscles in all the examined chicken breeds. In the thigh muscles, we have other pairs of significantly correlated genes: *GHR–MYF5*, *MEF2C–MYH1*, and *MYOG–MYF5*, suggesting clear differences in the myogenesis of different chick embryo muscle tissues. In terms of early growth and chick BW changes (Supplementary Tables S4, S5; Supplementary Figure S12), significant correlations were found for the two earlier measures (pre-incubation EW and post-hatch BW) as well as for the three postnatal BW measures (at 1, 14, and 28 days of life).

When considering the results of hierarchical breed clustering within this model, the broiler breeds BR and PRW constituted a close cluster, while another broiler breed, WC, was located to the side in all PCA plots and trees (Supplementary Figures S13, S14). The second large cluster consisted of egg-type LR, game UG and all dual purpose breeds, with one of the members of this cluster, YC, being located remotely from it. It is also worth noting that the preliminary transformation of the dataset into the Euclidean distance-based similarity matrix (similar to obtaining the Type IIIa data) reflected this clustering pattern even more clearly (Supplementary Figures S14B, C): WC stood away even further from BR and PRW, and the second cluster turned out to be very compact and crowded, with a very small separation of YC from it. In addition, judging from the trait clustering pattern (Supplementary Figure S13B), the used set of traits was clearly divided into two large clusters in terms of its contribution to interbreed differences. The first cluster included all DGE indicators (except for the correlated *MYOG* and *MYOD1* in the breast muscles), while the characteristics of early development and growth were in the second cluster.

Finally, based on this model that combines all the 21 traits studied in the eight breeds (including GR2wk and GR4wk), we observed again similar patterns in the clustering of breeds and features (Figure 8; Supplementary Figure S15). On the right side of the PCA plot for the distribution of the entire set of 21 traits accounted for in the eight breeds (Supplementary Figure S16), one can observe the main crowded core of DGE measures with several outliers, for example, *MEF2C* and *MYF5* (in the breast and thigh muscles), and *MYOG* (in the thigh muscles). From this core to the left side of the PCA graph, indicators of development and growth lined up almost along the same vector with increasing distance from NO and EW to BW28. Almost the same division into two large groups of features was noted on the hierarchical clustering plot in Figure 8.

Additionally, we calculated Spearman's correlation coefficients for pairwise comparison of eight breeds and the same 21 indicators (Supplementary Tables S6, S7; Supplementary Figure S17). Using

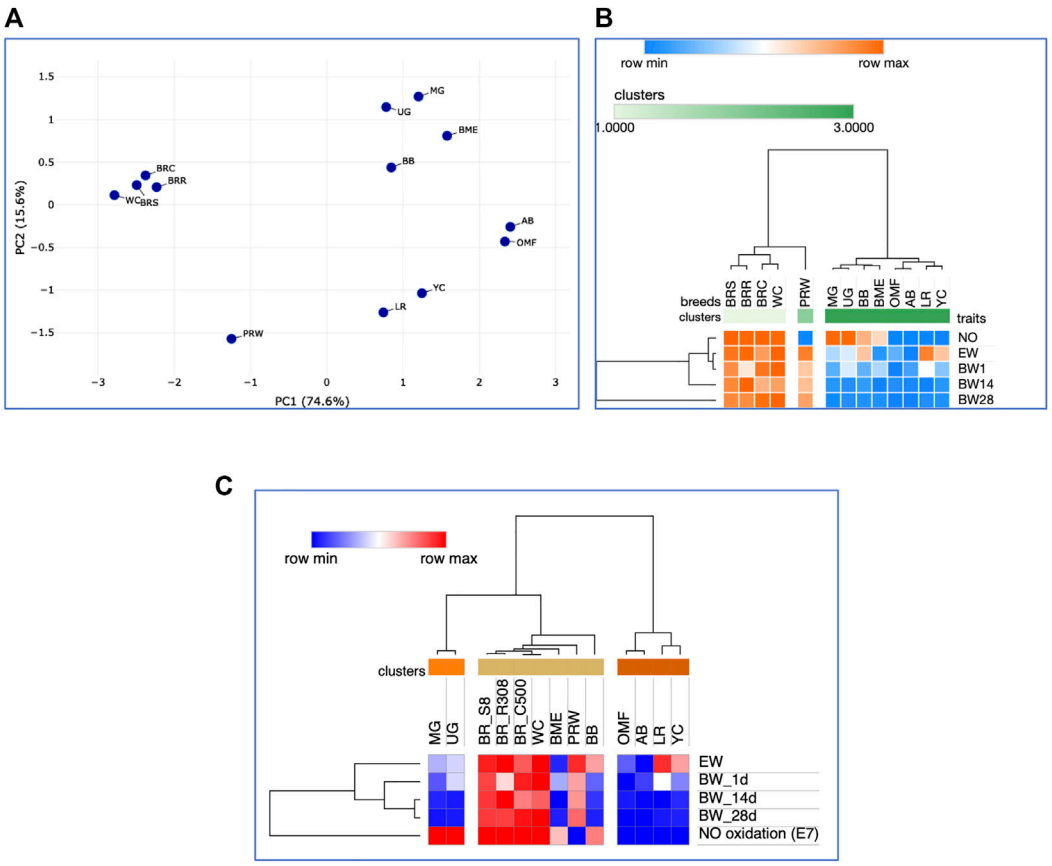


FIGURE 7
Analysis of the distribution of 13 the breeds by traits of E7 NO oxidation and postnatal growth, including EW and BW of chicks at three ages, performed in the Phantasus program (Zenkova et al., 2018). **(A)** PCA plot. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 74.6% and 15.6% of the total variance, respectively. *N* = 13 data points (breeds). **(B)** Heatmap and hierarchical clustering tree based on Euclidean distance metric (with the average option selected for the linkage method). **(C)** Heatmap and hierarchical clustering tree using one minus Pearson's correlation metric (with the average option as linkage method).

Spearman's correlation coefficient data, interbreed clustering patterns were tested on PCA and hierarchical clustering plots (Figure 9). These graphs showed the formation of a central core of breeds, composed of BR and their two parental forms WC and PRW, as well as the game UG. Egg LR, dual purpose OMF, and a subcluster of two dual purpose breeds BB and YC were located at a distance from this core and along differently directed vectors. Significant pairwise Spearman's correlations supported the previously found relationships between certain indicators of early myogenesis and postnatal growth (Supplementary Table S7; Supplementary Figure S17B). Thus, highly correlated DGE profiles of the *MSTN*, *GHR*, and *MEF2C* genes in the breast muscles were verified. In the thigh muscles, the DEG levels of *GHR* and *MYF5*, as well as *MYOG* and *MYOD1*, were positively correlated. The DGE indices of the *MSTN* gene in the breast and thigh muscles in different breeds had a high and significant correlation; this was also observed in the case of the *MEF2C* and *MYOG* genes. DGE of a few genes in the breast muscles positively correlated with that of other genes in the thigh muscles, e.g., *MYOG* in the breast muscles and *MYOG* in the thigh muscles. At the same time, when comparing DGE in pairs, some other genes negatively correlated with each other, in particular, *MYOD1* in the breast and *GHR* in the thigh muscles. All this contributed to the peculiar DGE profiles observed for the myogenesis associated genes studied. Mutual positive pairwise correlation between the early chick growth indicators (EW, BW1, BW14, BW28, and GR2wk) in various breeds was also confirmed.

4 Discussion

4.1 Embryogenesis, postnatal growth, and DGE of myogenesis associated genes

The present study suggested consistent DGE patterns of the *GHR* and *MSTN* genes, as well as *MYH1*. In certain breeds, however, myogenesis associated genes worked differently in the thigh muscles than in the breast muscles as evidenced by slightly different breed clustering patterns revealed by PCA and hierarchical clustering. Using various analytical approaches (e.g., Figure 1) these different effects of the myogenesis associated genes' functioning depending on the type of muscle were also demonstrated. As an example, there were a few reports describing the *MYOG* and *MYF5* genes expressed in concert (Auradé et al., 1994; Conerly et al., 2016), as we observed in the case of correlated DGE of these genes in the thigh muscles.

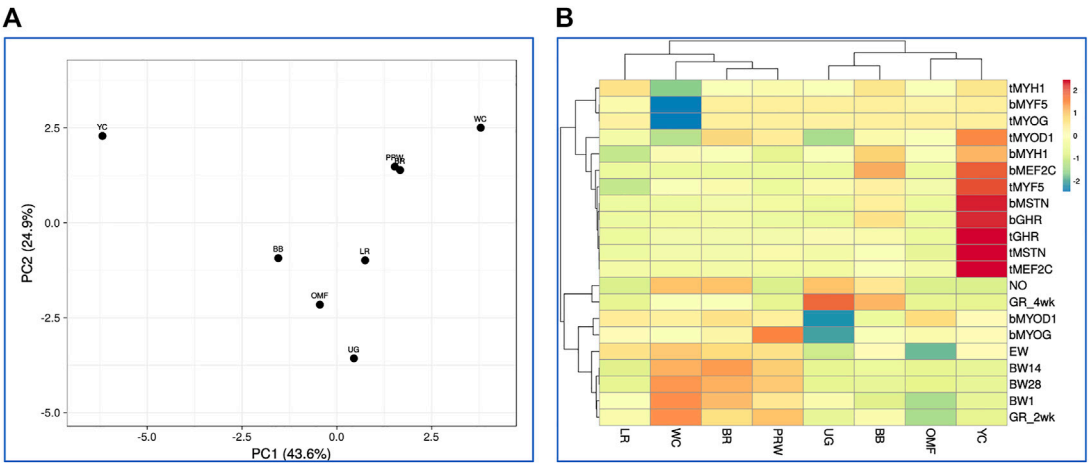


FIGURE 8
Distribution of the eight breeds based on the analysis of relationships between 21 traits (DGE of myogenesis associated genes and NO metabolism in embryos, as well as indicators of early chick growth) as performed in the ClustVis program (Metsalu and Vilo, 2015). **(A)** PCA plot. Unit variance scaling was applied to rows; singular value decomposition with imputation was used to calculate principal components. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 43.6% and 24.9% of the total variance, respectively. *N* = 8 data points (breeds). **(B)** Heatmap and clustering trees using Euclidean distances (with the average option selected as linkage method).

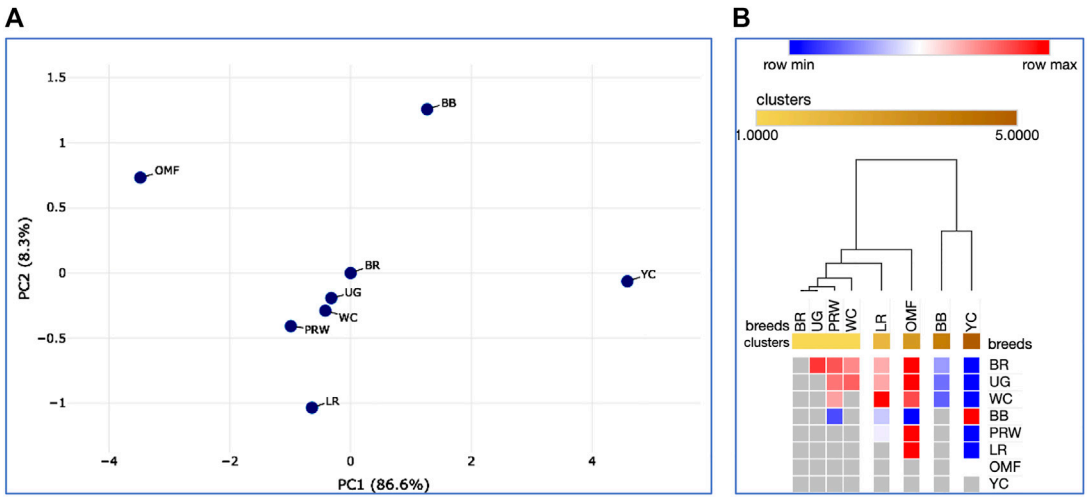


FIGURE 9
Analysis of the distribution of the eight breeds for 21 traits (DGE of myogenesis associated genes and NO metabolism in embryos, as well as indicators of early chick growth) performed in the Phantasus program (Zenkova et al., 2018). **(A)** PCA plot. *X* and *Y*-axes show principal component 1 (PC1) and principal component 2 (PC2) that explain 86.6% and 8.3% of the total variance, respectively. *N* = 8 data points (breeds). **(B)** Heatmap and hierarchical clustering tree based on Euclidean distance metric (with the average option as linkage method).

In the primary processing of DGE data reflecting features of the synthesis of mRNA molecules, it is important to develop solutions for reliable DEG identification. Genes are considered to be differentially expressed if they satisfy the *p*-value test and the FC test. If we take into account that the FC value is understood as a multiplicity factor, operations on it should be performed appropriately. In the present study, just such transformations were carried out. To establish DEG, they were ranked based on their FC (Mutch et al., 2002). A number of generally accepted procedures can be helpful in searching for DEG. When analyzing microarrays or RNA-Seq data for thousands of genes, the first step should include removal (filtering) of genes with a very low number in all libraries. There are both biological and statistical reasons for this (Chen et al., 2016). Thus, truncation, filtering, and transformation of data are primary tools in searching for DEG. The main purpose of these data manipulations is to narrow the search for genes of interest. In the current study only focused on the seven myogenesis associated genes, the task was not so much to search for DEG as there was a common goal—to isolate the characteristic indicators of early myogenesis in various chicken breeds created by divergent selection and belonging to one or another utility type.

As has already been established, embryonic metabolism can be divided into three major phases (Spiridonov et al., 2017). The first phase, or the embryonic period, begins in the oviduct and lasts in chickens up to E8. At this time, temporary embryonic organs are already functioning; nutrients are supplied from the yolk; and breathing occurs through the blood vessels of the yolk sac and, at the end of this phase, additionally through the vessels of the allantois. In the second "pre-fetal" period (from E9 to E14), nutrition occurs from yolk and then intra-intestinally *via* amniotic fluid; respiration—with the help of allantois; excretion of metabolic products through the mesonephros. The third "fetal" period (from E14 to E20) is characterized by the most rapid growth of the permanent organs of the embryo, nutrition with protein dissolved in the amniotic fluid, excretion of uric acid through the metanephros, and allantoic respiration (Spiridonov et al., 2017). The extraction of nutrients from the protein and yolk is largely commensurate with the body growth until the embryo completion by E14. Since E15, the metabolic profile of the embryo muscles changes. From E15 to E19, the chick embryo prepares for hatching by increasing the relative mass of the liver and muscles, by elevating the concentration of protein in the muscles, and by accumulating glucose and glycogen in it (Pulikanti et al., 2010). The hourglass model of embryo development suggests that the middle, or phylotypic, stage of embryonic development, when the body plan characteristic of this type of animal is laid down, have an increased evolutionary conservatism compared to the early and late stages. In addition, it turned out that genes that work at the middle development stages are characterized by increased multifunctionality: Many of them perform various functions at different development stages and in different parts of the body (Irie and Kuratani, 2014; Furusawa and Irie 2020).

Genes that control the middle stages of development are characterized by increased pleiotropy (multifunctionality). Many of them are involved not only in the rapid morphogenetic processes of the phylotypic stage of development, but also in other processes at other stages. Apparently, these genes were involved more often in the course of evolution than others to perform novel functions, e.g., when old regulatory genes can acquire new functions. Multifunctional genes operating at the phylotypic stage are so important for the normal development of an organism, that the system of their DGE regulation gained increased noise resilience in the course of evolution (Hu et al., 2017). Metabolic pathways are dependently linked to each other and share intermediate metabolite substrates, so they require precise homeostatic control. The amount and type of substrates available to the embryo trigger the production of hormones, which in turn control DGE of genes for enzymes that regulate the flows in these pathways. From about E5, the chorioallantoic membrane begins to develop, and from E8 it becomes the main means of oxygen uptake. In this study, we turned our attention to the turning points that occur at E7, when the creation of chorioallantois occurs and there is a change in the process of respiration from limited to high oxygen consumption (Tullett and Deeming, 1982; Reijrink et al., 2008). The second turning point of embryonic development is E14. The breast muscle growth hormone (GH) and its receptor (GHR) in the third period of embryonic development gradually begin to be expressed and reach its peak 48 h before hatching (data obtained in turkeys; de Oliveira et al., 2013). It is known that turkey embryos reach full body size by the pipping stage (3 days before hatching), and an increase in muscle mass is a consequence of increased tissue hydration (Vleck, 1991). *GHR* is one of genes considered when calculating indices for the breast and thigh muscles. As we found, there was a positive correlation between the DGE indices for *GHR*, *MSTN*, and *MEF2C* in the breast muscles at the level of .93 (p-value <.001).

In terms of GR patterns, we discovered a relationship between the increase in BW and utility type change from egg to meat type. These data also showed that the game breed, UG, is not characterized by the same body muscularity and GR as compared to the meat breeds, meaning that:

1. Commercial meat-type breeds and BR crosses have the highest rates of BW growth as a result of long-term artificial selection for meat traits.
2. The game breed was not subject to such selection. For any game breed, the most important are fighting qualities with a fairly light BW, which ensures the mobility of the bird in cock fights arranged in the past. Therefore, the BW growth in game chickens is more consistent with that in egg-type breeds.

4.2 The mechanism of interrelation of embryonic NO oxidation and post-hatch GR

From the results of our study (Supplementary Table S2), it follows that in E7 embryos (more precisely, in their muscle tissues) in almost all meat breeds and crosses, a high degree of NO oxidation to nitrate takes place. At the same time, NO oxidation to nitrate in embryos of egg-type breeds is minor. Selection for increased meat performance within the same breed (Andalusian Blue) resulted in an elevation in the oxidation degree of NO synthesized during embryogenesis in the BMET breed. Because BMET is a product of the Andalusian Blue chickens selected for meat traits, the degree of NO oxidation in the BMET embryos was ~62%, while in the Andalusian Blue, like in all egg-type breeds, it was marginal (~2%; Supplementary Table S2). Also, GR4wk in the BMET breed was significantly higher than that in the Andalusian Blue breed (p < .05; Supplementary Table S2).

An even greater difference in GR was observed, for example, for the BRS–LR pair (Supplementary Table S2). Broilers are produced by crossing male and female grandparent stocks, which in turn are also the result of crossing certain lines of WC and PRW breeds, respectively. Supplementary Table S2 shows that the three broiler crosses and their male grandparent stock breed (WC) were characterized by almost complete NO oxidation in the embryos (~97%–98%). Conversely, in the female grandparent stock breed (PRW), NO is practically not oxidized. All these data, on the one hand, indicate that the oxidation degree of embryonic NO is genetically determined. However, the nature of its inheritance suggests that the intensity of NO oxidation is determined not by any specific gene but, apparently, by the combination of several DEG, although we were unable to detect significant association of NO oxidation with the seven myogenesis associated genes tested. It is also known that the oxidation degree of embryonic NO does not depend on the incubation conditions, as well as the age and maintenance conditions of female breeders (Titov et al., 2012; Titov et al., 2018).

Based on the data obtained, we can hypothesize that NO is involved in specific processes of avian embryogenesis. First of all, the fact that in BR embryos most of the deposited NO (~90%) is oxidized to nitrate suggests that the high concentrations of deposited NO we observe in the amnion of egg-type breeds are not essential, at

least for supporting vital NO-dependent processes. As previously shown (Titov et al., 2018; Dolgorukova et al., 2020), the NO oxidation to nitrate occurs in the embryo tissues and mainly in muscle tissue. NO oxidation is practically absent in the liver and intestines. That is, we can assume that this oxidation is somehow associated with the development of muscle tissue. As for the role of NO deposited in the embryo of egg-type breeds, it may play a role as a pool in case of activation of NO oxidation processes, which can occur in any embryo.

Our analysis of the obtained data (Supplementary Table S2) shows that a high rate of NO oxidation is generally typical for meat-type and game chickens. It should be borne in mind that meat-type breeds are those that are profitable for raising birds for meat production, considering that they grow relatively quickly, and the gain in BW is ensured by relatively low feed costs. Note that the yield of gutted carcasses in broilers is only 5% higher than that in egg-type chickens. Therefore, the main feature for meat-type poultry is a rapid increase in BW. The breeds, lines and crosses listed in Supplementary Table S2 that had a high degree of NO oxidation in the embryos were also characterized by a more intensive growth of BW as compared to those with a lower degree of oxidation (Royter et al., 2005; Vinnikova and Titov, 2008). From our data (Supplementary Table S2), it also follows that NOD compounds are initially accumulated in the embryos. Starting from a certain point, these compounds begin to oxidize to nitrate. In egg-type embryos, NO oxidation is practically negligible. The key moments and processes of embryonic myogenesis are related to the fact that myotomes are laid down in E2 and E3 chick embryos, and the proliferation of myoblasts occurs up to E14. The process of NO oxidation in meat-type chick embryos occurs throughout the entire embryogenesis. Histological studies did not reveal any qualitative differences in the development of muscle tissues in BR vs. egg-type embryos characterized by respectively high and low rates of embryonic NO oxidation (Titov et al., 2018). It can be assumed that some factors associated with NO oxidation appear at E2 or E5 and this, apparently, is genetically determined and mediated by DGE of many genes involved in muscle development (Titov et al., 2018; Titov et al., 2020b).

Cazzato et al. (2014) studied DGE of some important myogenesis associated genes at the earliest stages of embryogenesis and showed the effect of NOSI and NOD on DGE. According to our data, inhibition of NO synthase at the initial stage of embryogenesis by 80% did not significantly affect the postembryonic GR (Titov et al., 2018; Dolgorukova et al., 2020). Of interest is the difference not in the intensity of NO synthesis, which is approximately the same in all embryos of the same species, but in the degree of its oxidation that differs many times in fast-growing vs. slow-growing chickens. Therefore, our data suggest that: 1) NO oxidation degree is genetically determined and inherited; 2) it is determined by DGE of not one but many genes; and 3) there are chances for activation of NO oxidation in all avian embryos. It can also be assumed that it is not NO that primarily affects DGE, but DGE affects NO oxidation (Titov et al., 2020b). In other words, the process of NO oxidation can be triggered by internal genetic factors (Titov et al., 2018) and partially by external factors (Figure 10).

It can be hypothesized that NO oxidation is catalyzed by some heme-containing protein, similar to the process observed in the interaction of NO with oxyhemoglobin (Herold, 1999). What role NO oxidation itself plays is still not completely clear, however, this process can serve as a biochemical marker of the breed characteristics related to development (myogenesis) both in the embryonic and postembryonic periods (Dolgorukova et al., 2020). Being initiated at the beginning of embryo development, NO oxidation continues throughout the entire embryonic stages suggesting that, under the influence of genetically determined (and external) factors, a population of cells is formed, within which the oxidation process occurs. The exact mechanism of this intracellular interaction of different pathways may be associated with specific biochemical signaling networks (Bhalla and Iyengar, 1999) and should be studied further.

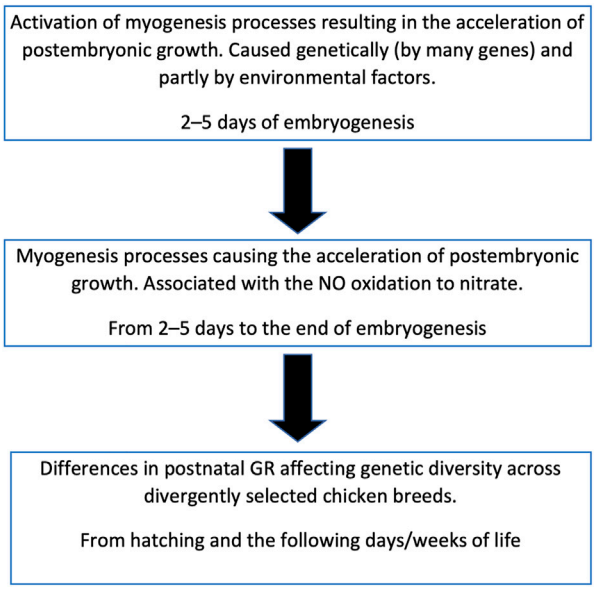


FIGURE 10
Scheme of relationship between the embryonic myogenesis processes, postnatal growth and genetic diversity in chicken breeds.

By confirming, and elucidating details of, some important phenomena in the chick embryo development, our findings expand the basic knowledge of how the early myogenesis genes work and how NO oxidation is involved in this process in various chicken breeds. Further investigation of these genetic signatures may have their practical significance as useful markers for the genetic breeding and genomic selection of chickens.

5 Conclusion

In the present study, we established that signatures of genetic diversity in divergently selected chicken breeds can be already traced at early developmental stages and be reflected in differences in embryonic myogenesis, NO metabolism, and postnatal growth patterns. Myogenesis associated genes were expressed in a coordinated manner, showing peculiar DGE and co-expression patterns depending on the type of muscle tissue under consideration (breast vs. thigh) and the type of divergent selection and utility to which this or that breed belonged. The coordinated ("accord") expression patterns of the genes *MSTN*, *GHR*, and *MEFC2* in the breast and thigh muscles served as genetic diversity markers among the breeds under study. Additionally, related expression vectors for the *MYOG* and *MYOD1* genes in the breast muscles as well as *MYOG* and *MYF5* in the thigh muscles were discovered. It was

demonstrated that the main part of NO synthesized in the avian embryo plays a specific role and can be accumulated in tissues as part of NOD compounds or be oxidized to nitrate. Being a biochemical marker for breed-specific characteristics that determine the rate of muscle mass growth (Titov et al., 2020b; Dolgorukova et al., 2020), NO oxidation correlated differently with early myogenesis in divergently selected breeds of different utility types: in BR embryos, NO was oxidized to nitrate by ~90%, while in egg-type embryos, oxidation was minor. It is assumed that the degree of NO oxidation in embryonic tissues is genetically determined (Titov et al., 2018) and caused not by a specific gene but, apparently, by a combination of many DEG associated with the NO oxidation to nitrate. Postembryonic growth patterns correlated with features of early muscle development and NO metabolism were generally consistent with, and accurately captured, evolutionary history of divergently selected chicken breed types reflecting their existing genetic diversity.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The animal study was reviewed and approved by the Animal Welfare Committee of the FSBEI HE "Moscow State Academy of Veterinary Medicine and Biotechnology—MVA named after K. I. Skryabin" and Federal Scientific Center "All-Russian Poultry Research and Technological Institute" of the Russian Academy of Sciences.

Author contributions

VT, IN, and MR conceived the idea and outline of the manuscript. VT, EB, NV, and MR provided the methodology substantiation. MK, OM, and AD carried out the lab investigation. EB, NV, and MR were responsible for the software support. EB, NV, and MR conducted the formal analysis. IK, IN, MK, OM, and AD provided the resources support. VT and IN curated the data. VT, EB, and MR wrote an original draft of the manuscript. MR and DG prepared, reviewed and proofread the final version of the manuscript. VT, EB, and MR were responsible for the visualization. IK, VT, and DG provided the supervision. IK was responsible for the project administration. IK, VT, and IN secured the funding acquisition. All authors reviewed and approved the manuscript for submission.

Funding

This study performed at the FSBEI HE "Moscow State Academy of Veterinary Medicine and Biotechnology—MVA named after K. I. Skryabin" was financially supported by the Russian Science Foundation (Grant No. 22-16-00009).

Acknowledgments

We thank all the research assistants who contributed to this work.

Conflict of interest

Author EB was employed by the company BIOTROF+ Ltd.

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

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.2022.1092242/full#supplementary-material

References

Abdelmanova, A. S., Dotsev, A. V., Romanov, M. N., Stanishevskaya, O. I., Gladyr, E. A., Rodionov, A. N., et al. (2021). Unveiling comparative genomic trajectories of selection and key candidate genes in egg-type Russian White and meat-type White Cornish chickens. *Biology* 10 (9), 876. doi:10.3390/biology10090876

Anderson, J. E. (2000). A role for nitric oxide in muscle repair: Nitric oxide–mediated activation of muscle satellite cells. *Mol. Biol. Cell.* 11 (5), 1859–1874. doi:10.1091/mbc.11.5.1859

Auradé, F., Pinset, C., Chafey, P., Gros, F., and Montarras, D. (1994). Myf5, MyoD, myogenin and MRF4 myogenic derivatives of the embryonic mesenchymal cell line C3H10T1/2 exhibit the same adult muscle phenotype. *Differentiation* 55 (3), 185–192. doi:10.1046/j.1432-0436.1994.5530185.x

Bernini, F., Bagnato, A., Marelli, S. P., Zaniboni, L., Cerolini, S., and Strillacci, M. G. (2021). Genetic diversity and identification of homozygosity-rich genomic regions in seven Italian heritage Turkey (*Meleagris gallopavo*) breeds. *Genes.* 12 (9), 1342. doi:10.3390/genes12091342

Bhalla, U. S., and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* 283 (5400), 381–387. doi:10.1126/science.283.5400.381

Boc, A., Diallo, A. B., and Makarenkov, V. (2012). T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* 40 (W1), W573–W579. doi:10.1093/nar/gks485

Bogolyubsky, S. I. (1991). *Poultry breeding*. Moscow, USSR: Agropromizdat.

Cazzato, D., Assi, E., Moscheni, C., Brunelli, S., De Palma, C., Cervia, D., et al. (2014). Nitric oxide drives embryonic myogenesis in chicken through the upregulation of myogenic differentiation factors. *Exp. Cell. Res.* 320 (2), 269–280. doi:10.1016/j.yexcr.2013.11.006

Chen, Y., Lun, A. T., and Smyth, G. K. (2016). From reads to genes to pathways: Differential expression analysis of RNA-seq experiments using rsubread and the edgeR quasi-likelihood pipeline. *F1000Res.* 5, 1438. doi:10.12688/f1000research.8987.2

Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M., and Tapscott, S. J. (2016). Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation. *Dev. Cell.* 36 (4), 375–385. doi:10.1016/j.devcel.2016.01.021

de Oliveira, J. E., Druyan, S., Uni, Z., Ashwell, C. M., and Ferket, P. R. (2013). Metabolic profiling of late-term Turkey embryos by microarrays. *Poult. Sci.* 92 (4), 1011–1028. doi:10.3382/ps.2012-02354

Dementieva, N. V., Mitrofanova, O. V., Dysin, A. P., Kudinov, A. A., Stanishevskaya, O. I., Larkina, T. A., et al. (2021). Assessing the effects of rare alleles and linkage disequilibrium on estimates of genetic diversity in the chicken populations. *Animal* 15 (3), 100171. doi:10.1016/j.animal.2021.100171

Dimmeler, S., Haendeler, J., Nehls, M., and Zeiher, A. M. (1997). Suppression of apoptosis by nitric oxide via inhibition of interleukin-1beta-converting enzyme (ICE)-like and cysteine protease protein (CPP)-32-like proteases. *J. Exp. Med.* 185 (4), 601–607. doi:10.1084/jem.185.4.601

Dolgorukova, A. M., Titov, V. Y., Kochish, I. I., Fisinin, V. I., Nikonov, I. N., Kosenko, O. V., et al. (2020). The embryonic metabolism of nitric oxide and its interrelation with postembryonic development in chicken (Gallus gallus domesticus L.) and quails (*Coturnix coturnix* L.). *Sel'skokhozyaistvennaya Biol. Agric. Biol.* 55 (4), 794–803. doi:10.15389/agrobiology.2020.4.794eng

Furusawa, C., and Irie, N. (2020). Toward understanding of evolutionary constraints: Experimental and theoretical approaches. *Biophys. Rev.* 12 (5), 1155–1161. doi:10.1007/s12551-020-00708-2

Herold, S. (1999). Mechanistic studies of the oxidation of pyridoxalated hemoglobin polyoxyethylene conjugate by nitrogen monoxide. *Arch. Biochem. Biophys.* 372 (2), 393–398. doi:10.1006/abbi.1999.1534

Hickok, J. R., Sahni, S., Shen, H., Arvind, A., Antoniou, C., Fung, L. W., et al. (2011). Dinitrosyliron complexes are the most abundant nitric oxide-derived cellular adduct: Biological parameters of assembly and disappearance. *Free Radic. Biol. Med.* 51 (8), 1558–1566. doi:10.1016/j.freeradbiomed.2011.06.030

Hu, H., Uesaka, M., Guo, S., Shimai, K., Lu, T. M., Li, F., et al. (2017). Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol. Evol.* 1 (11), 1722–1730. doi:10.1038/s41559-017-0318-0

Huang, X., Zhang, J., He, D., Zhang, X., Zhong, F., Li, W., et al. (2016). Genetic diversity and population structure of indigenous chicken breeds in South China. *Front. Agric. Sci. Eng.* 3 (2), 97. doi:10.15302/J-FASE-2016102

Imangulov, Sh. A., Egorov, I. A., Okolelova, T. M., and Tishenkov, A. N. (2013). *Methodology for conducting scientific and industrial research on feeding poultry: Recommendations*. Editors V. I. Fisinin and S. Posad (Russia: VNITIP).

Irie, N., and Kuratani, S. (2014). The developmental hourglass model: A predictor of the basic body plan? *Development* 141 (24), 4649–4655. doi:10.1242/dev.107318

Kanakachari, M., Ashwini, R., Chatterjee, R., and Bhattacharya, T. (2022). Embryonic transcriptome unravels mechanisms and pathways underlying embryonic development with respect to muscle growth, egg production, and plumage formation in native and broiler chickens. *Front. Genet.* 13, 990849. doi:10.3389/fgene.2022.990849

Kassambara, A., and Mundt, F. (2017). Factoextra: Extract and visualize the results of multivariate data analyses. Version 1.0. Available at: https://cran.r-project.org/web/packages/factoextra/index.html (Accessed on November 3, 2022).

Kim, Y. M., Chung, H. T., Simmons, R. L., and Billiar, T. R. (2000). Cellular non-heme iron content is a determinant of nitric oxide-mediated apoptosis, necrosis, and caspase inhibition. *J. Biol. Chem.* 275 (15), 10954–10961. doi:10.1074/jbc.275.15.10954

Larkina, T. A., Barkova, O. Y., Peglivanyan, G. K., Mitrofanova, O. V., Dementieva, N. V., Stanishevskaya, O. I., et al. (2021). Evolutionary subdivision of domestic chickens: Implications for local breeds as assessed by phenotype and genotype in comparison to commercial and fancy breeds. *Agriculture* 11 (10), 914. doi:10.3390/agriculture11100914

Li, Y., Wang, Y., Willems, E., Willemsen, H., Franssens, L., Buyse, J., et al. (2016). In ovo L-arginine supplementation stimulates myoblast differentiation but negatively affects muscle development of broiler chicken after hatching. *J. Anim. Physiol. Anim. Nutr.* 100 (1), 167–177. doi:10.1111/jpn.12299

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Long, J. H., Lira, V. A., Soltow, Q. A., Betters, J. L., Sellman, J. E., and Criswell, D. S. (2006). Arginine supplementation induces myoblast fusion via augmentation of nitric oxide production. *J. Muscle Res. Cell. Motil.* 27 (8), 577–584. doi:10.1007/s10974-006-9078-1

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., et al. (2021). Package 'cluster'. Version 2.1.2 Available at: https://cran.r-project.org/web/packages/cluster/cluster.pdf (Accessed on November 3, 2022).

Metsalu, T., and Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43 (W1), W566–W570. doi:10.1093/nar/gkv468

Moiseyeva, I. G., Romanov, M. N., Nikiforov, A. A., Sevastyanova, A. A., and Semyenova, S. K. (2003). Evolutionary relationships of red jungle fowl and chicken breeds. *Genet. Sel. Evol.* 35 (4), 403–423. doi:10.1186/1297-9686-35-5-403

Mutch, D. M., Berger, A., Mansourian, R., Rytz, A., and Roberts, M. A. (2002). The limit fold change model: A practical approach for selecting differentially expressed genes from microarray data. *BMC Bioinform* 3, 17. doi:10.1186/1471-2105-3-17

Pedersen, T. L. (2021). *ggplot2*. RDocumentation. Available at: https://www.rdocumentation.org/packages/ggplot2/versions/3.3.5 (Accessed on November 3, 2022).

Pulikanti, R., Peebles, E. D., Keirs, R. W., Bennett, L. W., Keralapurath, M. M., and Gerard, P. D. (2010). Pipping muscle and liver metabolic profile changes and relationships in broiler embryos on days 15 and 19 of incubation. *Poult. Sci.* 89 (5), 860–865. doi:10.3382/ps.2009-00531

Reijrink, I., Meijerhof, R., Kemp, B., and Van Den Brand, H. (2008). The chicken embryo and its micro environment during egg storage and early incubation. *Worlds Poult. Sci. J.* 64 (4), 581–598. doi:10.1017/S0043933908000214

Romanov, M. N., Dementyeva, N. V., Terletsky, V. P., Plemyashov, K. V., Stanishevskaya, O. I., Kudinov, A. A., et al. (2017). Applying SNP array Technology to assess genetic diversity in Russian gene pool of chickens" in Proceedings of the International Plant and Animal Genome XXV Conference, 14. San Diego, CA, USASan Diego, CA, USA: Scherago International.

Romanov, M. N., Larkina, T. A., Barkova, O. Yu., Peglivanyan, G. K., Mitrofanova, O. V., Dementieva, N. V., et al. (2021). "Comparative analysis of phenotypic traits in various breeds representing the world poultry gene pool," in *Materials of the 3rd international scientific and practical conference on molecular genetic technologies for analysis of gene expression related to animal productivity and disease resistance* (Moscow, RussiaMoscow, Russia: Sel'skokhozyaistvennye tekhnologii), 52–63. doi:10.18720/SPBPU/2/z21-43

Romanov, M. N., and Weigend, S. (1999). "Genetic diversity in chicken populations based on microsatellite markers," in *Proceedings of the conference from jay lush to genomics: Visions for animal breeding and genetics, ames, IA, USA, 16–18 may 1999*. Editors J. C. M. Dekkers, S. J. Lamont, and M. F. Rothschild (Ames IA USA: Iowa State University, Department of Animal Science).

Romanov, M. N., and Weigend, S. (2001). Using RAPD markers for assessment of genetic diversity in chickens. *Arch. Geflugelkd.* 65 (4), 145.

Rössig, L., Fichtlscherer, B., Breitschopf, K., Haendeler, J., Zeiher, A. M., Mülsch, A., et al. (1999). Nitric oxide inhibits caspase-3 by S-nitrosation *in vivo*. *J. Biol. Chem.* 274 (11), 6823–6826. doi:10.1074/jbc.274.11.6823

RStudio Team (2016). *RStudio: Integrated development for R. RStudio*. Boston, MA, USA: RStudio, Inc. Available at: http://www.rstudio.com/ (Accessed on November 3, 2022).

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3 (6), 1101–1108. doi:10.1038/nprot.2008.73

Severina, I. S., Bussygina, O. G., Pyatakova, N. V., Malenkova, I. V., and Vanin, A. F. (2003). Activation of soluble guanylate cyclase by NO donors—S-Nitrosothiols, and dinitrosyl-iron complexes with thiol-containing ligands. *Nitric Oxide* 8 (3), 155–163. doi:10.1016/s1089-8603(03)00002-8

Socco, S., Bovee, R. C., Palczewski, M. B., Hickok, J. R., and Thomas, D. D. (2017). Epigenetics: The third pillar of nitric oxide signaling. *Pharmacol. Res.* 121, 52–58. doi:10.1016/j.phrs.2017.04.011

Spiridonov, I. P., Maltsev, A. B., and Dymkov, A. B. (2017). "Incubation of poultry eggs from A to Z: An encyclopedic referencedictionary," in *Scientific institution*. Omsk: Feder, 201.

Stamler, J. S., and Meissner, G. (2001). Physiology of nitric oxide in skeletal muscle. *Physiol. Rev.* 81 (1), 209–237. doi:10.1152/physrev.2001.81.1.209

Stamler, J. S., Singel, D. J., and Loscalzo, J. (1992). Biochemistry of nitric oxide and its redox-activated forms. *Science* 258 (5090), 1898–1902. doi:10.1126/science.1281928

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22 (12), 1540–1542. doi:10.1093/bioinformatics/btl117

Tarpey, M. M., Wink, D. A., and Grisham, M. B. (2004). Methods for detection of reactive metabolites of oxygen and nitrogen: *In vitro* and *in vivo* considerations. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 286 (3), R431–R444. doi:10.1152/ajpregu.00361.2003

Thompson, S., Romanov, M. N., and Griffin, D. K. (2021). "Study of animal myosins in a comparative genomic aspect," in *Materials of the 3rd international scientific and practical conference on molecular genetic technologies for analysis of gene expression related to animal productivity and disease resistance* (Moscow, RussiaMoscow, Russia: Sel'skokhozyaistvennye tekhnologii), 2021, 444–449. doi:10.18720/SPBPU/2/z21-43

Tirone, M., Conti, V., Manenti, F., Nicolosi, P. A., D'Orlando, C., Azzoni, E., et al. (2016). Nitric oxide donor molsidomine positively modulates myogenic differentiation of embryonic endothelial progenitors. *PLoS One* 11 (10), e0164893. doi:10.1371/journal.pone.0164893

Titov, V., Dolgorukova, A., Khasanova, L., Kochish, I., and Korenyuga, M. (2021). Nitric oxide (NO) and arginine as factors for increasing poultry meat productivity. *KnE Life Sci.* 6 (3), 622–631. doi:10.18502/kls.v0i0.8998

Titov, V. Y., Dolgorukova, A. M., Fisinin, V. I., Borkhunova, E. N., Kondratov, G. V., Slesarenko, N. A., et al. (2018). The role of nitric oxide (NO) in the body growth rate of birds. *Worlds Poult. Sci. J.* 74 (4), 675–686. doi:10.1017/S0043933918000661

Titov, V. Y., Dolgorukova, A. M., Vertiprakhov, V. G., Ivanova, A. V., Osipov, A. N., Slesarenko, N. A., et al. (2020a). Synthesis and metabolism of nitric oxide (NO) in chicken embryos and in the blood of adult chicken. *Bull. Exp. Biol. Med.* 168 (3), 321–325. doi:10.1007/s10517-020-04700-4

Titov, V. Y., Kosenko, O. V., Starkova, E. S., Kondratov, G. V., Borkhunova, E. N., Petrov, V. A., et al. (2016). Enzymatic sensor detects some forms of nitric oxide donors undetectable by other methods in living tissues. *Bull. Exp. Biol. Med.* 162 (1), 107–110. doi:10.1007/s10517-016-3557-1

Titov, V. Y., and Osipov, A. N. (2017). Nitrite and nitroso compounds can serve as specific catalase inhibitors. *Redox Rep.* 22 (2), 91–97. doi:10.1080/13510002.2016.1168589

Titov, V. Yu (2011). The enzymatic technologies open new possibilities for studying nitric oxide (NO) metabolism in living systems. *Curr. Enzym. Inhib.* 7 (1), 56–70. doi:10.2174/157340811795713774

Titov, V. Yu., Kochish, I. I., Nikonov, I. N., Korenyuga, M. V., Myasnikova, O. V., Kuvanov, T. K., et al. (2020b). "[Genetic markers of meat performance in poultry," in *Materials of the 2nd international scientific and practical conference on molecular genetic technologies for analysis of gene expression related to animal productivity and disease resistance* (Moscow, RussiaMoscow, Russia: Sel'skokhozyaistvennye tekhnologii), 25136. doi:10.18720/SPBPU/2/k20-5

Titov, V. Yu., Vinnikova, E. Z., Akimova, N. S., and Fisinin, V. I. (2012). Nitric oxide (NO) in bird embryogenesis: Physiological role and ability of practical use. *Worlds Poult. Sci. J.* 68 (1), 83–96. doi:10.1017/S0043933912000098

Tullett, S. G., and Deeming, D. C. (1982). The relationship between eggshell porosity and oxygen consumption of the embryo in the domestic fowl. *Comp. Biochem. Physiol. A Comp. Physiol.* 72 (3), 529–533. doi:10.1016/0300-9629(82)90118-9

Ulibarri, J. A., Mozdziak, P. E., Schultz, E., Cook, C., and Best, T. M. (1999). Nitric oxide donors, sodium nitroprusside and S-nitroso-N-acetylpencillamine, stimulate myoblast proliferation *in vitro*. *Vitro Cell. Dev. Biol. Anim.* 35 (4), 215–218. doi:10.1007/s11626-999-0029-1

Vanin, A. F., Borodulin, R. R., and Mikoyan, V. D. (2017). Dinitrosyl iron complexes with natural thiol-containing ligands in aqueous solutions: Synthesis and some physico-chemical characteristics (a methodological review). *Nitric Oxide* 66, 1–9. doi:10.1016/j.niox.2017.02.005

Vanin, A. F. (2009). Dinitrosyl iron complexes with thiolate ligands: Physico-chemistry, biochemistry and physiology. *Nitric Oxide* 21 (1), 1–13. doi:10.1016/j.niox.2009.03.005

Vasudevan, D., Bovee, R. C., and Thomas, D. D. (2016). Nitric oxide, the new architect of epigenetic landscapes. *Nitric Oxide* 59, 54–62. doi:10.1016/j.niox.2016.08.002

Vinnikova, E. Z., and Titov, V. Yu. (2008). Determination of phenotypically unexpressed forms of the ostrich. *Ptitsevod. Poult. Farming* 12, 33.

Vleck, D. (1991). "Water economy and solute regulation," in *Egg incubation: Its effects on embryo development in birds and reptiles* (New York, NY, USA: Cambridge University Press), 252.

Wei, T., and Simko, V. (2021). R package 'corrplot': Visualization of a correlation matrix. Version 0.90 Available at: https://github.com/taiyun/corrplot (Accessed on November 3, 2022).

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2021). *ggplot2: create elegant data visualisations using the grammar of graphics*. The Comprehensive R Archive Network (CRAN); Institute for Statistics and Mathematics, Vienna University of Economics and Business. Version 3.3.5 Available at: https://cran.r-project.org/web/packages/ggplot2/index.html (Accessed on November 3, 2022).

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY, USA: Springer-Verlag. doi:10.1007/978-0-387-98141-3

Zenkova, D., Kamenev, V., Sablina, R., Artyomov, M., and Sergushichev, A. (2018). Phantasus: Visual and interactive gene expression analysis. Available at: https://ctlab.itmo.ru/phantasus (Accessed on November 3, 2022).doi:10.18129/B9.bioc.phantasus

Zhao, Q., Hautamaki, V., and Fränti, P. (2008). "Knee point detection in BIC for detecting the number of clusters," in *Lect Notes Comput Sci*. Editors J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders (Berlin/Heidelberg, Germany: Springer), 5259, 664.

Zhou, J., and Brüne, B. (2005). NO and transcriptional regulation: From signaling to death. *Toxicology* 208 (2), 223–233. doi:10.1016/j.tox.2004.11.021

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 January 2023
DOI 10.3389/fgene.2022.1081175

Check for updates

OPEN ACCESS

EDITED BY
Adnan Khan,
Agricultural Genomics Institute at Shenzhen (CAAS), China

REVIEWED BY
Sangang He,
Xinjiang Academy of Animal Science, China
Herman Revelo,
Fundación Universitaria San Martín, Colombia

*CORRESPONDENCE
Brenda M. Murdoch,
bmurdoch@uidaho.edu

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

RECEIVED 26 October 2022
ACCEPTED 20 December 2022
PUBLISHED 23 January 2023

CITATION
Becker GM, Woods JL, Schauer CS, Stewart WC and Murdoch BM (2023), Genetic association of wool quality characteristics in United States Rambouillet sheep.
Front. Genet. 13:1081175.
doi: 10.3389/fgene.2022.1081175

COPYRIGHT
© 2023 Becker, Woods, Schauer, Stewart and Murdoch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genetic association of wool quality characteristics in United States Rambouillet sheep

Gabrielle M. Becker[1], Julia L. Woods[1], Christopher S. Schauer[2], Whit C. Stewart[3] and Brenda M. Murdoch[1]*

[1]Department of Animal, Veterinary and Food Science, University of Idaho, Moscow, ID, United States, [2]Hettinger Research Extension Center, North Dakota State University, Hettinger, ND, United States, [3]Department of Animal Science, University of Wyoming, Laramie, WY, United States

Introduction: Fine wool production is an important source of revenue, accounting for up to 13% of total revenue in extensively managed wool sheep production systems of the United States. The Rambouillet are a predominant breed that excels in wool quality characteristics. Understanding the genetic basis of wool quality characteristics would aid in the development of genomic breeding strategies to facilitate genetic improvement.

Methods: Wool characteristics and DNA were collected for rams enrolled in the North Dakota State University and University of Wyoming annual central performance ram tests over a three-year period (2019–2021, *N* = 313). The relationships of wool quality characteristics including grease fleece weight adjusted 365 days (wt. 365 adj.), clean fleece wt. 365 adj., staple length 365 adj., average fiber diameter, face wool cover, amount of skin wrinkles and belly wool were evaluated through genome-wide association studies (GWAS), Pearson correlation and ANOVA.

Results: The GWAS identified four genome-wide significant genetic markers (*p*-value <1.19e-06) and five chromosome-wide significant markers (*p*-value <1.13e-05) on chromosomes 1, 2, 4, 15, and 19. Significant markers were associated with genes notable for relevant wool biological functions, including the gene *ABCC8* which codes for SUR1, an ATP-sensitive potassium channel known to affect hair growth and 60S ribosomal protein L17-like, previously found to be expressed during follicle formation. The strongest Pearson correlation coefficients were identified between clean fleece wt. 365 adj. and grease fleece wt. 365 adj. (r = 0.83) and between clean fleece wt. 365 adj. and staple length 365 adj. (r = 0.53). Additionally, clean fleece wt. 365 adj. was correlated with final body weight (r = 0.35) and scrotal circumference (r = 0.16). Staple length 365 adj. (*p*-value = 5e-04), average fiber diameter (*p*-value = .0053) and clean fleece wt. 365 adj. (*p*-value = .014) were significantly associated with belly wool score.

Discussion: The results of this study provide important insight into the relationships between wool quality characteristics and report specific markers that Rambouillet sheep producers may use to help inform selection and breeding decisions for improved wool quality.

KEYWORDS

60S ribosomal protein L17-like, ABCC8, central performance ram test, GWAS, sheep production

Abbreviations: 365 adj., adjusted to 365 days; CO, Colorado; MT, Montana; ND, North Dakota; SD, South Dakota; WY, Wyoming.

1 Introduction

Rambouillet are a predominant United States breed in extensive and semi-extensive production systems. This breed is commonly used in arid and semi-arid rangeland systems as a dual-purpose breed excelling in both fine wool and meat products (Lupton et al., 2007; Burton et al., 2015). Wool production is an important source of strategically timed revenue for sheep producers and fine wool receipts account for up to 13% of total revenue from sheep production in the United States (Liver Marketing Information Center, 2016; Murphy et al., 2019). Wool quality characteristics are well studied in Rambouillet and other fine-wool breeds, yet the genomic mechanisms underpinning these traits are still poorly defined and underutilized in genomic breeding strategies.

Wool quality is driven by clean fleece weight and fiber diameter (Khan et al., 2012) but many characteristics can contribute to the overall economic worth. Central performance ram tests have been developed as a way to systematically evaluate ram growth and performance traits under comparable environmental conditions with demonstration and outreach value for sheep producers (Shelton et al., 1954; Burton et al., 2015). Ram tests are held annually at North Dakota State University (NDSU) and the University of Wyoming (UWY) to evaluate Rambouillet and other wool breeds enrolled by local sheep producers.

Much progress has been made in sheep production through the identification and utilization of genetic markers for disease susceptibility risk or carrier identification (Westaway et al., 1994; Cockett et al., 1999), reproduction traits (Ivanova et al., 2021) and carcass and milk traits (Clop et al., 2006; Selvaggi et al., 2015). Wool quality characteristics have been previously estimated to be moderately to highly heritable, indicating that trait variation is greatly influenced through genetic effects and progress may be made through genomic selectin (Bromley et al., 2000; Burton et al., 2015). Despite such promising heritability estimates, few validated markers exist for use with Rambouillet genomic breeding strategies.

The aim of this study was to utilize data collected during NDSU and UWY central performance ram tests over a three-year period to characterize relationships between traits and with genomic single nucleotide polymorphism (SNP) markers. Pearson correlation and analysis of variance (ANOVA) testing were conducted with wool characteristics grease fleece weight adjusted to 365 days (wt. 365 adj.), clean fleece wt. 365 adj., staple length 365 adj., average fiber diameter, face wool score, skin wrinkle and belly wool scores and production traits including initial and final weights, 140-day average daily gain (ADG) and scrotal circumference. Wool traits were evaluated in individual GWAS with 50 k genotype data to identify markers for use in genomic breeding strategies.

2 Materials and methods

2.1 Ram test protocols

Ram lambs 7 ± 3 months of age from regional (WY, ND, SD, MT, CO) seedstock producers were brought to the University of Wyoming—Laramie Research and Extension Center (Laramie, WY; 41°17′ N, −105°40′ W) or North Dakota State University—Hettinger Research and Extension Center (Hettinger, ND; 46°01′ N, −102°65' W). Initial body weights were measured and animals were managed as one cohort. Rams were provided *ad libitum* textured diets (15%–17% crude protein, dry matter basis; 68%–73% total digestible nutrients dry matter basis) for 140 days in a dry-lot management system.

Rams were shorn after a 7–10-day acclimation period and once again at the conclusion of the 140-day feeding period. Upon conclusion of the performance test, scrotal circumference was obtained and wool staple length was measured on shoulder, side and britch, the three measurements averaged, and adjusted from 140-day to 365-day lengths in accordance with the standard practice of the National Sheep Improvement Program (NSIP) for this trait (Wilson and Morrical, 1991). This was calculated by dividing the average staple length by 140, to calculate staple length/day, and then multiplying by 365.

The presence of belly-type wool (belly wool) was scored from 1 to 4. In brief, belly wool is that which grows on the ventral region of the sheep and is characterized as uneven, tender in tensile strength, and compressed in staple length. Phenotypic selection pressure against this "belly wool" fiber type extending beyond the ventral portions of the sheep has been employed to avoid the resultant reduction in overall wool quality (Lupton et al., 2007; Naidoo et al., 2016). Thus, a subjective 1 to 4 scoring system was assigned in the fleece to rams at the end of the test period where: 1 = belly wool restricted to ventral portion, 2 = belly wool restricted to lower 1/3rd of side of fleece, 3 = belly wool extending from 1/3rd to ½ of the side of fleece, 4 = wool extending above ½ of the side of the fleece. Rams were scored linearly between these thresholds with a score of 1 being the minimum and a score of 4 being the highest possible (e.g., a ram with belly wool extending midway between the ventral portion and 1/3 of the side would be scored 1.5).

Similarly, a subjective 1 to 4 scoring system was assigned for face cover where: 1 = no wool cover over top of the head and on the side of muzzle, nor between eyes and ears, 2 = minimal wool cover over top of the head and on side of muzzle, and between eyes and ears, 3 = moderate wool cover over top of the head and on side of muzzle, and between eyes and ears, 4 = heavy wool cover over top of the head and on side of muzzle, and between eyes and ears. A skin wrinkle score was assessed once wool was shorn, where: 1 = no observable wrinkles on body surface 2 = minimal observable wrinkles on body surface 3 = moderate observable wrinkles on body and 4 = heavy wrinkles on body surface.

At shearing, whole fleeces were weighed and then individually cored in a custom-built apparatus (Gerbers of Montana, Inc., Great Falls, MT) consisting of 16 coring tubes (2.2 cm in diameter) that were plunged into and retracted from compacted fleeces by hydraulic cylinders. Cores were split into duplicate 25-g sub-samples for each animal to determine average laboratory scoured yield (American Society for Testing and Materials, 1990) from which clean fleece weight (CFW) was also estimated (Grease Fleece Weight x LSY). Grease and clean fleece weights were adjusted from 140-day to 365-day lengths in the same manner described for staple length (Wilson and Morrical, 1991). A single washed core subsample was analyzed on an Optical-based Fibre Diameter Analyser 2000 (OFDA; BSC Electronics Pty. Ltd., Attadale, Western Australia) to quantify average fiber diameter (A-FD) (IWTO, 2013).

2.2 Statistical evaluation of wool characteristics

Wool characteristics analyzed included grease fleece weight adjusted to 365 days (wt. 365 adj.) and clean fleece wt. 365 adj. (pounds), average fiber diameter (micron), staple length 365 adj. (inches), face wool score and skin wrinkle score as continuous variables and belly wool score as a

TABLE 1 *p*-values for location and year against wool characteristics. Wool characteristics were tested against location (NDSU or UWY) with *t*-test and year (2019, 2020, 2021) with ANOVA.

	Location (*t*-test)	Year (ANOVA)
Grease Fleece Wt. 365 adj.	7.23E-04*	7.21E-01
Clean Fleece Wt. 365 adj.	2.39E-03*	1.81E-03*
Staple Length 365 adj.	1.79E-10*	1.29E-01
Average Fiber Diameter	3.55E-01	4.93E-03*
Face Wool Score	1.23E-05*	1.27E-01
Skin Wrinkle Score	<2.2e-16*	4.29E-02*

*indicates significant *p*-values of covariates included as fixed effects in EMMAX GWAS for trait model.

categorical variable. All traits were tested for normality using the Shapiro Wilks test in R version 4.2.1 (R Core Team 2021); face and skin wrinkle scores were transformed using a log10 transformation to improve normality. Belly wool scores were grouped into the variable "belly wool category" with rams with no belly-type wool comprising category one (n = 273), rams with belly wool on less than one-third of the side comprising category two (n = 25) and rams with belly wool from one-third of the side to over one-half of the side comprising category three (n = 15). Wool characteristics were analyzed against other production traits including initial body weight and final body weight (pounds), 140-day average daily gain (140 days ADG) and scrotal circumference (centimeters).

The relationships between continuous wool quality characteristics and production traits were investigated with Pearson correlation to describe the strength and direction of linear correlation. One-way analysis of variance (ANOVA) testing was utilized to compare production trait and continuous wool quality trait means between belly wool categories. All ANOVA tests were further analyzed with *post hoc* Tukey HSD testing to compare *p*-values between categories (Abdi and Williams, 2010). Ram test location (North Dakota or Wyoming) was evaluated by Welch's two-sample *t*-test and ram test year (2019, 2020, 2021) was tested by ANOVA to determine significance for potential GWAS fixed effects (Table 1). Pearson correlation testing were conducted and visualized using the corrplot package in R (Wei and Simko, 2021). ANOVA and Tukey HSD were conducted with the rstatix package and visualized with ggplot2, ggpubr and patchwork in R (Kassambara 2020a; Kassambara 2020b; Pedersen 2020).

2.3 DNA genotyping and quality control

Ram DNA samples were extracted from either whole blood samples or ear tissue samples stored in tissue sampling units (TSU) collected by University of Wyoming or North Dakota State University personnel. DNA was isolated from blood at the University of Idaho using the phenol-chloroform method described previously (Sambrook et al., 1989) and TSU were provided to AgResearch for DNA extraction. Ram DNA samples were genotyped with either the Applied Biosystems™ Axiom™ Ovine Genotyping Array (50K) consisting of 51,572 single nucleotide polymorphism (SNP) markers (Thermo Fisher Scientific, catalog number 550898) or the AgResearch Sheep Genomics 60K SNP chip consisting of 68,848 SNP markers (GenomNZ, AgResearch, New Zealand). Duplicate markers designed for the same genomic position within a panel were filtered to retain the marker with the highest call rate (CR). Compatible markers were matched by marker name and genome position resulting in a consensus dataset of 44,431 markers in common between the genotype platforms (Davenport et al., 2020). Plink v1.9 was used to merge genotype array data and correct markers designed for opposite strands (Purcell et al., 2007; https://pngu.mgh.harvard.edu/purcell/plink/). Markers were filtered for quality control in the following order: non-autosomal markers (1,019 SNPs), markers with a call rate (CR) <90% (87 SNPs), markers with a minor allele frequency (MAF) <0.01 (1,407 SNPs) and markers with Hardy-Weinberg Equilibrium *p*-values <1e-50 (30 SNPs) were excluded, for a total of 41,888 high-quality autosomal SNPs retained for final analyses. All rams had a CR of 95% or greater.

2.4 Principal component analysis

Principal component analysis (PCA) was carried out to investigate population structure. Analysis was conducted with plinkv1.9 and visualized with the package ggplot2 in R (Purcell et al., 2007; https://pngu.mgh.harvard.edu/purcell/plink/; Kassambara 2020b; R Core Team 2021). Principal components were plotted PC1 (x-axis) *versus* PC2 (y-axis). Separate plots were generated for each continuous wool characteristic and rams were color-coded on a gradient scale to indicate their position within the trait distribution.

2.5 Genome-wide association studies

Continuous wool characteristics were evaluated in genome-wide association studies (GWAS) using the Efficient Mixed-Model Association eXpedited (EMMAX) in SNP and Variation Suite™ v8.9.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The EMMAX models estimated the proportion of variance explained (PVE) for each marker as previously described (Kang et al., 2010). Each trait was initially tested in additive, dominant and recessive inheritance models to identify the model of best fit to be carried through for final analysis. A genomic relationship matrix was fitted as a random effect to account for population structure and sample relatedness in each model (Kang et al., 2010). Ram test location and ram test year were fitted as fixed effects as warranted by *t*-test or ANOVA *p*-value for each trait (Table 1) and GWAS results were visualized with the CMplot package in R (Yin 2022). Genome-wide significance was determined by the Bonferroni threshold (*p*-values <1.

TABLE 2 Descriptive statistics of wool quality characteristics. Wool quality characteristics of the 313 study rams collected from NDSU and UWY central performance ram tests over 3 years.

	Grease fleece wt. 365 adj. (Lb.)	Clean fleece wt. 365 adj. (Lb.)	Staple length 365 adj. (in.)	Average fiber diameter (micron)	Face wool score	Skin wrinkle score
Average ±SD	20.79 ± 3.13	11.82 ± 2.04	5.08 ± 0.57	22.64 ± 1.51	1.30 ± 0.50	1.39 ± 0.45
Min	13.10	7.09	3.10	19.01	1.00	1.00
Median	20.70	11.63	5.09	22.56	1.00	1.25
Max	31.00	18.04	6.97	27.20	3.40	3.50
Range	17.90	10.95	3.87	8.19	2.40	2.50

SD, standard deviation.

19e-06) and a chromosome-wide significance threshold was determined by Bonferroni-adjustment of the number of markers on the largest chromosome (4,412 markers; p-values <1.13e-05). The trait distributions of significant markers were visualized using boxplot figures and significance was further evaluated through analysis of covariance (ANCOVA) and Tukey HSD testing in R with the rstatix, ggplot2, ggpubr, multcomp and patchwork packages (Hothorn et al., 2008; Abdi and Williams, 2010; Kassambara 2020a; Kassambara 2020b; Pedersen 2020; R Core Team 2021). Each ANCOVA model included the same covariate(s) as included in the corresponding GWAS model.

2.6 Genomic context of significant markers

The genomic contexts of significant markers were investigated using GenomeBrowser in NCBI (NCBI Resource Coordinators 2016) for the reference genome ARS-UI_Ramb_v2.0 (Davenport et al., 2022). For each genome-wide and chromosome-wide significant SNP, the reference sequence comprising 100 kb upstream and 100 kb downstream of the marker were evaluated. Markers positioned within a gene were further evaluated for predicted transcription factor binding site (TFBS) score differences between major and minor alleles. The online software FABIAN (Steinhaus et al., 2022) was utilized to test query sequences against detailed transcription factor flexible models (TFFM) compiled within the JASPAR 2022 database (Kahn et al., 2018). Query sequences were comprised of 11 bp, including the five nucleotides flanking the marker on the 5′ and 3′ side in the reference genome. Reference sequences were tested with the major allele as the "wild-type" sequence and the minor allele as "variant" sequence. Where applicable, the *Homo sapiens* ortholog of each associated gene was queried through ProteomeHD and STRING databases to identify potential interactions between genes/proteins implicated in the study results (Szklarczyk et al., 2015; Kustatscher et al., 2019).

3 Results

3.1 Statistical evaluation of wool characteristics

3.1.1 Pearson's correlation tests for wool characteristics and production traits

Descriptive statistics of wool characteristics are reported for the 313 rams (Table 2), by test location (Supplementary Table S1) and by test year (Supplementary Table S2). Relationships between continuous wool characteristics were evaluated using Pearson correlation tests. The strongest relationship was identified between clean fleece wt. 365 adj. and grease fleece wt. 365 adj. (r = 0.83; p-value = 3.22e-80) (Table 3). Grease fleece wt. 365 adj. was significantly correlated (p-value <.05) with all traits tested. Average fiber diameter was found to have significant positive correlations with both clean and grease fleece 365 adj. weights (r = 0.19; r = 0.24 and p-value = 6.27e-04; p-value = 1.55e-05) respectively, and clean fleece wt. 365 adj. had significant positive correlations with staple length 365 adj. (r = 0.53; p-value = 6.42e-24) and skin wrinkle score (r = 0.14; p-value = 1.40e-02). Skin wrinkle and face wool scores had a significant positive correlation (r = 0.26; p-value = 4.09e-06). Clean and grease fleece 365 adj. weights were significantly correlated with initial body weight (r = 0.23; r = 0.18), final body weight (r = 0.35; r = 0.43), 140 days ADG (r = 0.25; r = 0.39) and scrotal circumference (r = 0.16; r = 0.24).

3.1.2 Relationship of wool quality characteristics to presence of belly wool

Belly wool score was found to have significant relationships with initial body weight and final body weight, with rams in category three tending to have greater weights than rams in category one (Tukey HSD p-value ~0.01) (Figure 1). Significant relationships were identified with wool characteristics staple length 365 adj. (p-value = 5e-04), average fiber diameter (p-value = 0.5.3e-3) and clean fleece wt. 365 adj. (p-value = 1.4e-2). Post hoc Tukey HSD tests revealed that rams within belly wool category two had significantly longer staple length 365 adj. than category one (p-value = 2.93e-4) and category three (p-value = 3.47e-2) rams. Rams with belly wool scores in category three had significantly finer average fiber diameter than rams in category one (p-value = 1.54e-2), although there was no significant difference between rams in categories three and two or two and one. For clean fleece wt. 365 adj., rams in belly wool category two had significantly greater measurements than rams in category one (p-value = 1.05e-2) (Figure 2). The relationships between belly wool and grease fleece wt. 365 adj., face wool score, skin wrinkle score, scrotal circumference and 140 days ADG were also investigated and were not found to be significant.

3.2 Principal component analysis

Principal component analysis (PCA) was used to investigate the population structure as it related to wool quality characteristics. Plots were constructed with principal component 1 (PC1) on the x-axis and principal component 2 (PC2) on the y-axis. PC1 had an eigenvalue of

10.3389/fgene.2022.1081175

TABLE 3 Pearson correlation results for ram production and wool characteristics. Correlation coefficients (*r*) are reported on the upper diagonal and *p*-values are reported on the lower diagonal.

	Grease fleece wt. 365 adj.	Clean fleece wt. 365 adj.	Staple length 365 adj.	Face wool score	Skin wrinkle score	Average fiber diameter	SC	Initial weight	Final weight	140 days ADG
Grease Fleece Wt. 365 adj.	—	0.83*	0.41*	0.12*	0.28*	0.24*	0.24*	0.18*	0.43*	0.39*
Clean Fleece Wt. 365 adj.	3.22E-80	—	0.53*	0.09	0.14*	0.19*	0.16*	0.23*	0.35*	0.25*
Staple Length 365 adj.	5.97E-14	6.42E-24	—	0.08	0.16*	−0.04	0.03	0.05	0.28*	0.36*
Face Wool Score	3.41E-02	1.10E-01	1.41E-01	—	0.26*	−0.01	−0.13*	−0.05	−0.01	0.08
Skin Wrinkle Score	4.26E-07	1.40E-02	4.56E-03	4.09E-06	—	0.05	−0.09	−0.04	0.25*	0.45*
Average Fiber Diameter	1.55E-05	6.27E-04	4.42E-01	8.05E-01	3.94E-01	—	0.12*	0.20*	0.16*	0.01
SC	1.65E-05	3.95E-03	6.34E-01	1.93E-02	1.09E-01	3.81E-02	—	0.26*	0.42*	0.26*
Initial Weight	1.10E-03	3.64E-05	3.73E-01	3.59E-01	4.61E-01	3.39E-04	2.37E-06	—	0.66*	−0.05
Final Weight	2.77E-15	2.09E-10	3.50E-07	8.56E-01	8.13E-06	5.97E-03	1.13E-14	6.96E-40	—	0.71*
140 days ADG	4.95E-13	7.89E-06	4.27E-11	1.83E-01	3.91E-17	8.71E-01	3.18E-06	3.68E-01	4.51E-49	—

*indicates a significant *p*-value (<.05). Face wool score and skin wrinkle score were tested as log10 transformed data. 140 days ADG, average daily gain over 140 days; SC, scrotal circumference.

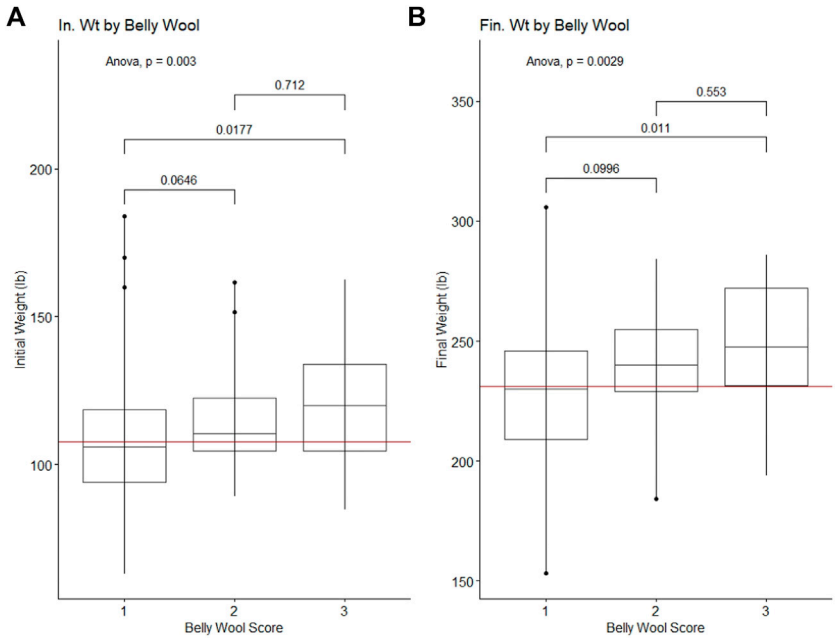


FIGURE 1
Significant ANOVA results and post-hoc Tukey HSD *p*-values for production traits tested against belly wool categories. **(A)** Initial on-test weight, **(B)** Final test weight. Horizontal red lines indicate the trait mean.

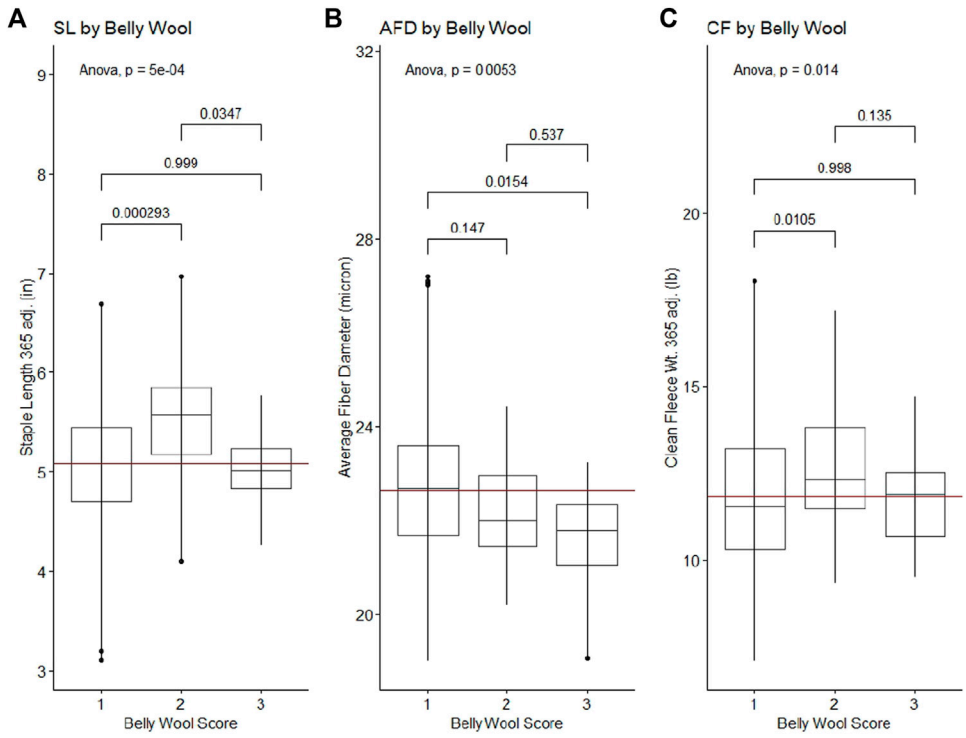


FIGURE 2
Significant ANOVA results and post-hoc Tukey HSD *p*-values for wool quality characteristics tested against belly wool categories. **(A)** Staple length 365 adj., **(B)** average fiber diameter, **(C)** clean fleece wt. 365 adj. Horizontal red lines indicate the trait mean.

8.57 and explained 11.17% of the total variance, PC2 had an eigenvalue of 6.91 and explained 9.00% of the total variance (Figure 3). There does not appear to be any specific clustering of phenotypically similar rams in the first or second PC for the wool traits examined. Color-coding of rams by wool quality characteristic distributions indicates these specific characteristics do not segregate with any particular genetic relationships.

3.3 Genome-wide association studies

Genome-wide association studies (GWAS) were conducted for each of the six continuous wool quality characteristics. Ram test location and/or ram test year were included as fixed effects for traits with significant (*p*-value <0.05) ANOVA or *t*-test results. The results of GWAS are displayed in a multi-trait Manhattan plot (Figure 4A) and individual quantile-quantile (QQ) plots (Figure 4B) and unadjusted *p*-values are reported (Table 4). Three SNPs on chromosome 1 reached genome-wide significance, including two SNPs associated with average fiber diameter and one SNP associated with clean fleece wt. 365 adj. Significant SNPs for average fiber diameter were identified in a dominant inheritance model (rs404487383 with *p*-value = 2.53e-07; rs406184307 with *p*-value = 5.11e-07) and were estimated to explain 8.25% and 7.85% of phenotypic variance. The significant SNP rs420943224 was found to be significant for clean fleece wt. 365 adj. by genome-wide threshold and for grease fleece wt. 365 adj. by chromosome-wide threshold in the corresponding additive inheritance models (rs420943224; *p*-value = 1.16e-06; *p*-value = 4.27e-06) with 7.40% and 6.60% proportion of variance explained (PVE), respectively. Two significant SNPs were identified on chromosome 15 for skin wrinkle score (rs402689377; additive) and staple length (OAR15_66653722.1; recessive) and three significant SNPs on chromosomes 2, 4, and 19 were associated with face wool score in a recessive inheritance model (OAR2_197807108.1; rs429550684; OAR19_14805437.1). The PVE for significant SNPs ranged from 6.13% to 8.25% and MAF ranged from 5.13% to 48.40%.

3.3.1 Marker validation through ANCOVA

The relationships between significant GWAS SNP genotypes and their associated wool quality characteristics were further evaluated through ANCOVA and Tukey HSD tests. The mean trait values for the alternate homozygous genotype, heterozygous genotype and reference homozygous genotype of each SNP are reported (Supplementary Table S3). Rams homozygous for the major allele (CC) at rs406184307 were found to have significantly lower average fiber diameter measurements than rams heterozygous (CT; *p*-value < 1e-04) or homozygous for the minor allele (TT; *p*-value = 2.35e-04) (Figure 5A). Presence of one or two copies of the C allele at rs420943224 had significantly greater mean clean fleece wt. 365 adj. than rams homozygous for the major allele (TT) with *p*-value = 3.49e-04 and 1.79e-02, respectively (Figure 5B). Boxplot figures for remaining significant SNPs are located in Supplementary Figures S1, S2.

3.4 Genomic context of significant markers

To investigate the genomic context of GWAS results, the reference genome sequence was evaluated for the presence of known or

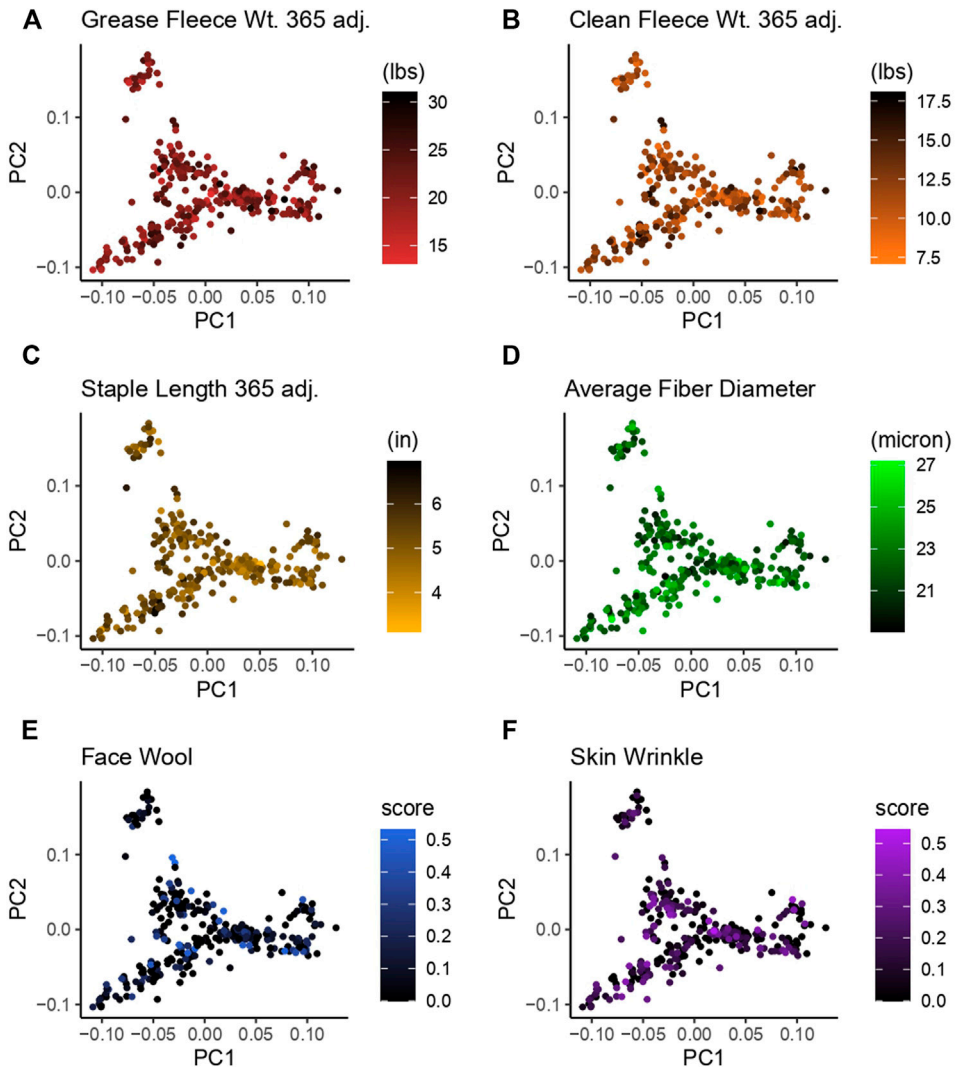


FIGURE 3
Principal component analysis (PCA) for 313 Rambouillet rams. Each panel represents PC1 plotted on the x-axis and PC2 plotted on the y-axis. Rams are color-coded based on their position within the trait distribution, with the most desirable end of the distribution represented by black. Each panel is color-coded low to high: **(A)** grease fleece wt. 365 adj., red to black; **(B)** clean fleece wt. 365 adj., orange to black; **(C)** staple length 365 adj., yellow to black; **(D)** average fiber diameter, black to green; **(E)** face wool score, black to blue, **(F)** skin wrinkle score, black purple. Face wool and skin wrinkle are colored based on log10 transformed data.

predicted genes (Table 4). The functional consequences of SNPs within genes were further investigated through TFBS prediction analysis. Reference genome sequence for the markers rs404487383 within 60S ribosomal protein L17-like (LOC121818710), rs402689377 within ATP binding cassette subfamily C member 8 (*ABCC8*) and OAR19_14805437.1 within unc-51 like Kinase 4 (*ULK4*) were queried for TFBS differences. Score difference between reference and alternate allele sequences of 0.3 or −0.3 or greater are recorded (Table 5). Four SOX family TFBS were predicted at rs404487383 and a YY2 and MZF1 TFBS were predicted at rs402689377. There were no TFBS with score difference of 3/-3 or greater predicted at SNP OAR19_14805437.1. Query of *Homo sapiens* ortholog genes through ProteomeHD and STRING databases revealed a known interaction between human proteins RPL17 and RPS19 with a co-regulation percentile score of 0.9998 (https://www.proteomehd.net/proteomehd/P18621/0.989988).

4 Discussion

To the authors' knowledge, this study represents the first GWAS conducted for wool quality characteristics of Rambouillet sheep. Genetic markers for wool quality traits have been previously identified through GWAS for other breeds of sheep, including Merino and Chinese fine-wool sheep (Wang et al., 2014; Zhao et al., 2021a; Bolormaa et al., 2021; Zhao B. et al., 2021), North-Caucasian sheep (Krivoruchko et al., 2022) and Baluchi sheep (Ebrahimi et al., 2017). Of note, the marker rs410503867 reported for association with super-elite rams (Krivoruchko et al., 2022) was positioned 2.3 Mb from a marker significant for face wool (rs429550684) in the current study. Additionally, Zhao et al. (2021a) reported markers within candidate genes *USP13* and *NLGN1* associated with staple length and positioned 1.8 Mb and 3.3 Mb, respectively, from markers identified for average fiber diameter (rs404487383; rs406184307) in the present study.

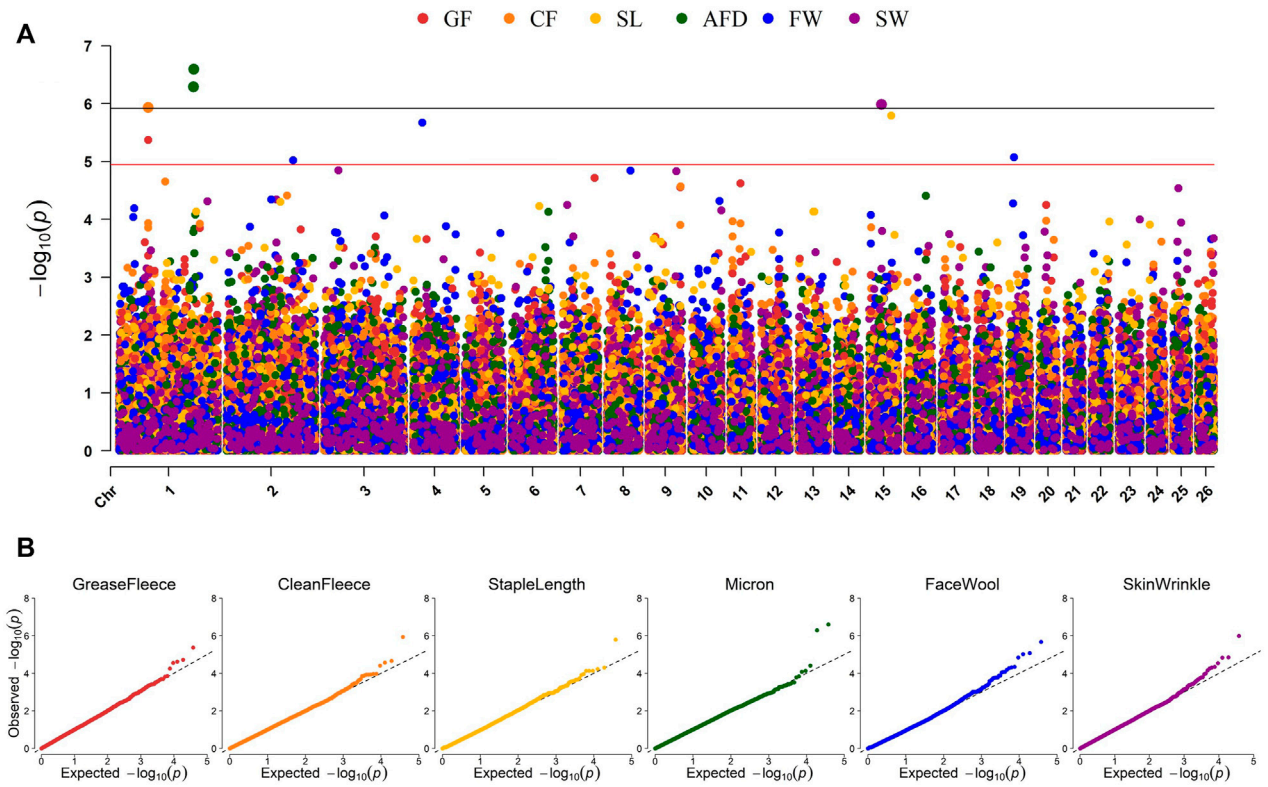


FIGURE 4
Manhattan and QQ plots representing EMMAX GWAS results for six continuous wool characteristics. **(A)** Manhattan plot representing the GWAS results of six wool traits. *p*-values are represented by: grease fleece wt. 365 adj., red; clean fleece wt. 365 adj., orange; staple length 365 adj., yellow; average fiber diameter, green; face wool score, blue; skin wrinkle score, violet. Genome-wide significance is given by *p*-values <1.19e-06 (black line) and chromosome-wide significance is given by *p*-values <1.13e-05 (red line). **(B)** Quantile-Quantile (QQ) plots for each GWAS displaying the expected *versus* observed−log10 (*p*-value).

Genetic markers for wool quality in sheep have also been suggested through candidate gene studies, including markers associated with genes *MTR* (Rong et al., 2015), *FST* (Ma et al., 2017), *DKK1* (Mu et al., 2017), *KIF16B* (Zhao et al., 2021c), *FGF5* (Zhao et al., 2021d) and keratin-associated proteins (Gong et al., 2016; Itenge, 2021). In candidate gene studies within the Rambouillet breed, several keratin intermediate filament (*KRT*) and keratin-associated protein (*KAP*) genes have been suggested for genomic selection (Mahajan et al., 2017a; Mahajan et al., 2017b; Mahajan et al., 2019; Singh et al., 2022). Despite this body of literature, there is still a need for robust genome-wide investigations for markers associated with wool characteristics of Rambouillet sheep.

The current study investigated a genome-wide distribution of SNP markers for significance against six wool quality characteristics. Of the eight significant SNPs identified, seven markers were located in proximity to at least one gene, and three of these markers were located within a gene. Genes containing significant SNPs have been previously associated with biological functions relevant to follicle growth (Figure 6). The marker rs402689377 associated with skin wrinkle score was located within the gene ATP binding cassette subfamily C member 8 (*ABCC8*) which codes for SUR1, an ATP-sensitive potassium channel known to affect fiber growth (Shorter et al., 2008). Although *ABCC8* has not been previously associated with wool quality in sheep, another potassium voltage-gated channel-related gene (*KCNIP4*) has been suggested to be related to sheep growth (Mohammadi et al., 2020). The marker OAR19_14805437.1 associated with face wool score was within unc-51 like Kinase 4 (*ULK4*). Overexpression of the gene *ULK4* has been shown to inhibit apoptosis (Luo and Yang 2022). Fiber producing follicles undergo a cycle including active growth, apoptosis-driven involution (catagen), shedding and resting in humans and mice (Botchkareva et al., 2006). The wool follicles of sheep are known to undergo similar cyclic activity with periods of catagen (Liu et al., 2013). The most significant SNP for average fiber diameter, rs404487383, was within 60S ribosomal protein L17-like. This gene was previously found to be one of the 50 most highly expressed genes within regenerating velvet skin of Red deer (Yang et al., 2016), suggesting a potential role in the sheep skin transcriptome.

Significant SNPs were located in regions containing biologically relevant genes. The marker rs429550684 was significant for face wool score and was located downstream of 40S ribosomal protein S19-like (LOC121818710). The gene similar to ribosomal protein S19 (LOC364797) was identified within the sparse and wavy hair (*swh*) locus of rats (Kuramoto et al., 2005), although annotation for this gene has since been withdrawn from the *Rattus norvegicus* assembly as it was not predicted in a later annotation (NCBI Gene ID: 364797). The *Homo sapiens* orthologs of 40S ribosomal protein S19-like and 60S ribosomal protein L17-like have known protein interactions, suggesting a potential for similar interaction of these proteins in sheep. The *swh* locus is known to be associated with follicle

TABLE 4 Results of GWAS for wool quality characteristics. Each trait was tested individually in an EMMAX model and significant markers (genome-wide, *p*-values <1.19e-06; chromosome-wide, *p*-values <1.13e-05) are reported.

Marker ID	rs number	Chr: Position (bp)	Trait	Model	COV	*p*-value	MAF (%)	PVE (%)	Genomic context
OAR1_224418361.1	rs404487383	1:210,457,046	AFD	D	Y	2.53e-07	36.22	8.25	Within 60S ribosomal protein L17-like (LOC121818710) (Yang et al., 2016)
OAR1_224016330.1	rs406184307	1:210,061,545	AFD	D	Y	5.11e-07	48.40	7.85	Downstream of LOC121816904 (lncRNA)
s29455.1	rs402689377	15:34,799,858	SW	A	P, Y	1.03e-06	25.40	7.47	Intronic, ATP binding cassette subfamily C member 8 (ABCC8) (Shorter et al., 2008)
OAR1_86433231.1	rs420943224	1:81,908,905	CF	A	P, Y	1.16e-06	5.13	7.40	Upstream of U6 spliceosomal RNA (LOC114110993) (Hilcenko et al., 2013)
OAR15_66653722.1	—	15:61,931,743	SL	R	P	1.62e-06	16.77	7.16	Downstream of WT1 (Wagner et al., 2008), upstream of LOC105602333, upstream of translation machinery-associated protein 7-like (LOC114118447), upstream of LOC114118448 (lncRNA)
OAR4_26881691.1	rs429550684	4:26,484,846	FW	R	P	2.14e-06	44.63	7.00	Upstream of LOC121819390 (lncRNA), downstream of 40S ribosomal protein S19-like (LOC101106000) (Kuramoto et al., 2005)
OAR1_86433231.1	rs420943224	1:81,908,905	GF	A	P	4.27e-06	5.13	6.60	Upstream of U6 spliceosomal RNA (LOC114110993) (Hilcenko et al., 2013)
OAR19_14805437.1	—	19:14,279,682	FW	R	P	8.49e-06	33.65	6.20	Intronic, ULK4 (unc-51 like Kinase 4) (Luo and Yang 2022)
OAR2_197807108.1	—	2:187,691,398	FW	R	P	9.58e-06	9.97	6.13	Intergenic

Chr, chromosome; BP, base pair position; AFD, average fiber diameter; SW, skin wrinkle score; CF, clean fleece wt. 365 adj.; SL, staple length 365 adj.; FW, face wool score; GF, grease fleece wt. 365 adj.; A, additive inheritance; D, dominant inheritance; R, recessive inheritance; COV, covariate; Y, year; P, place; MAF, minor allele frequency; PVE, proportion of variance explained.

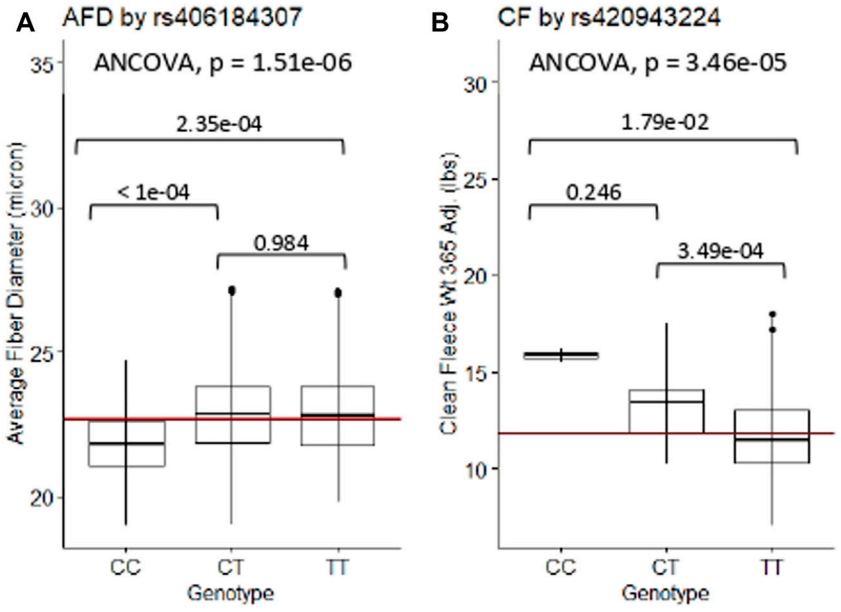


FIGURE 5
ANCOVA and post-hoc Tukey HSD test results of selected significant SNPs. **(A)** Average fiber diameter against SNP rs406184307, and **(B)** Clean fleece wt. 365 adj. against significant SNP rs420943224. The red horizontal lines indicate the trait means. AFD, average fiber diameter; CF, clean fleece wt. 365 adj.

hypoplasia, as well as impaired development of the sebaceous glands and mammary glands (Kuramoto et al., 2005). The marker rs420943224 associated with clean fleece wt. 365 adj. and grease fleece wt. 365 adj. was located upstream of U6 spliceosomal RNA (LOC114110993). A spliceosomal U6 small nuclear RNA has been previously indicated in poikiloderma with neutropenia (Hilcenko

TABLE 5 Predicted TFBS for SNPs located within genes. Query sequences were analyzed with the major allele as "wild type" and the minor allele as "variant" sequence. The score depicts the difference of wild type *versus* variant predictions.

Marker ID	Predicted TFBS	Score	Query
rs404487383	SOX10	−0.91	TCTTT[T/A]GTTGC
	SOX2	−0.50	
	SOX3	−0.49	
	SOX17	−0.39	
rs402689377	YY2	−0.30	GAAAG[G/C] CCGAG
	MZF1	0.41	
OAR19_14805437.1	—	—	AGTGA[T/A]TCTGG

et al., 2013). Finally, the marker OAR15_66653722.1 associated with staple length 365 adj. was located downstream of *WT1*, which has been identified in fibroblast cells that can induce and support hair growth (Wagner et al., 2008). The proximity of significant SNPs to these genes suggests the possibility for linkage disequilibrium with untested causative markers, or the possibility for identified SNPs to be positioned within transcriptional regulatory elements. Further work is needed to elucidate the implications of these associations.

Prediction analysis for TFBS suggested binding motifs for SOX2, SOX3, SOX10 and/or SOX17 may have less specific binding abilities between the alternate and reference alleles at rs404487383. SOX2 is expressed in mesenchymal cells during skin development (Sarkar and Hochedlinger, 2013), and both SOX2 and SOX3 are involved in the development of inner-ear hair cells in zebrafish (Gou et al., 2018). The gene SOX10 has been shown to play an important role in the development of vestibular hair cells in the pig (Qi et al., 2022). Two potential TFBS were predicted with matrix score changes between SNP alternate and reference alleles at s29455.1, a marker associated with skin wrinkle scores. The TF MZF1 has been shown to diminish the expression of the gene *PADI1* in human keratinocyte cells (Dong et al., 2008). These *in silico* analyses suggest functional ramifications of variant alleles associated with wool quality characteristics.

The trait correlations reported in the current study largely agree with previously published work for Rambouillet sheep. Clean fleece wt. 365 adj. is a component of grease fleece wt. 365 adj., which is reflected in their robust correlation in this study and previous (Vesely et al., 1970). Clean fleece wt. 365 adj. was significantly correlated with both staple length 365 adj. and average fiber diameter, which agrees with relationships previously reported for clean fleece and fiber diameter (Lupton et al., 2002; Hanford et al., 2005; Mahajan et al., 2018). This study found no significant correlation between staple length 365 adj. and average fiber diameter, although staple length and fiber diameter (wool grade) have been previously reported to be either favorably or unfavorably related (Lupton et al., 2002; Hanford et al., 2005). There was a significant positive relationship between clean fleece wt. 365 adj. and scrotal circumference, which was similar to observations made in Merino rams and in other central performance ram tests (Duguma et al., 2002; Lupton et al., 2002). The presence of significant correlations between wool quality characteristics indicates that progress in one trait may either positively or negatively impact progress in another trait; for instance, gains in clean fleece weight may come at the expense of fiber diameter.

Similar associations as those reported between belly wool category and wool quality characteristics in the present study have been previously noted in Merino sheep (Naidoo et al., 2016). Genetic correlation between wool quality and 'creeping belly' scores, representing the extent of belly-type wool on the side of the sheep, were reported to be −0.55 ± 0.27. This relationship was interpreted as sheep with more acceptable wool quality also tended to have less acceptable scores for creeping belly. Creeping belly has been reported to be correlated with body weight in Afrino sheep (Snyman and Olivier, 2002), and importantly, has been reported to have an unfavorable correlation with reproduction (Snyman and Olivier, 2002). The current study suggested potential positive phenotypic associations

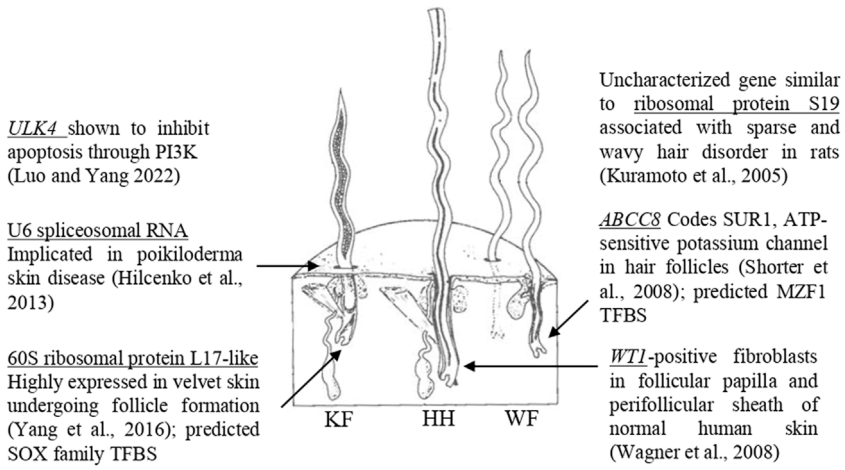


FIGURE 6
Genes with biological functions relevant to follicular growth. Significant SNPs identified in GWAS are located within or near genes with known biological roles relevant to skin and follicular growth, including *ULK4* (Luo and Yang 2022), U6 spliceosomal RNA (Hilcenko et al., 2013), 60S ribosomal protein L17-like (Yang et al., 2016), ribosomal protein S19 (Kuramoto et al., 2005), *ABCC8* (Shorter et al., 2008) and *WT1* (Wagner et al., 2008). The significant SNPs within 60S ribosomal protein L17-like and *ABCC8* had predicted TFBS score differences between reference and alternate alleles. The figure illustrates the three types of fibers which comprise sheep's wool, adapted from Bradford and Fitzhugh, 1983. KF, kemp fiber; HH, heterotype hair; WF, wool fiber.

between belly score category two and staple length 365 adj., average fiber diameter, clean fleece wt. 365 adj. and ram initial and final body weights, although further evaluation is needed to understand other important associations with belly wool in Rambouillet sheep.

The PCA investigated in this study indicated an overall lack of segregation of genetically similar animals with any of the specific wool quality characteristics tested by PC1 or PC2, suggesting an opportunity for genetic progress for wool quality characteristics. Previously reported heritabilities suggest that genetics are a major factor in wool quality phenotypes and genetic improvements should be possible even in the short term (Medrado et al., 2021). This study suggests specific genetic markers that may be used in marker-assisted selection programs for wool quality in Rambouillet sheep to make gains in economically important traits such as average fiber diameter and clean fleece weight. Interpretation of the current study is somewhat limited by sample size, as some genotype categories (such as CC at rs420943224) have as few as two rams observed. Improving sample sizes in underrepresented genotypes would improve statistical power and overall understanding of genotypic relationships with traits.

Data availability statement

The datasets have been deposited to EVA repository: https://urldefense.com/v3/__https://www.ebi.ac.uk/eva/?eva-study=PRJEB58836__;!!JYXjzlvb!gEHSLvJ5dvR-1g4UW6svxSQlbb8v0CZDShYTgf1HgGPLj3rxwSZ0_u0CaQFKVS5cKDCOr7UFqZw7zvCuQZ4qjmD-Q$ Accession Details are: Project: PRJEB58836, Analyses: ERZ15609617.

Ethics statement

The animal study was reviewed and approved by North Dakota State University Institute for Animal Care and Use Committee (# 20210012). Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

CS, WS, and BM conceived the study, WS and CS and University staff oversaw animal care, handling and data collection, GB, JW, CS, WS, and BM conducted research work and contributed to manuscript writing and editing. All authors read and agree to the manuscript.

Funding

This research was funded by the Idaho Global Entrepreneurial Mission and USDA-NIFA-IDA1566 and Hatch-Multistate "Increased Efficiency of Sheep Production" project accession no. 1025808 from the USDA National Institute of Food and Agriculture.

Acknowledgments

The authors would like to thank the sheep producers who contributed to the success of this study and acknowledge AgResearch and the Animal Genomics team for access to the AgResearch Sheep Genomics 60K SNP chip.

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.2022.1081175/full#supplementary-material

References

Abdi, H., and Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encycl. Res. Des.* 3 (1), 1–5.

American Society for Testing and Materials (ASTM) (1990). *Annual book of ASTM standards. Standard test method D584. Wool content of raw wool-laboratory scale*. Philadelphia, PA: ASTM, 193–197. Sec. 7, Vol. 07.01.

Bolormaa, S., Swan, A. A., Stothard, P., Khansefid, M., Moghaddar, N., Duijvesteijn, N., et al. (2021). A conditional multi-trait sequence GWAS discovers pleiotropic candidate genes and variants for sheep wool, skin wrinkle and breech cover traits. *Genet. Sel. Evol.* 53 (1), 58–14. doi:10.1186/s12711-021-00651-0

Botchkareva, N. V., Ahluwalia, G., and Shander, D. (2006). Apoptosis in the hair follicle. *J. Investigative Dermatology* 126 (2), 258–264. doi:10.1038/sj.jid.5700007

Bradford, G. E., and Fitzhugh, H. A. (1983). "Hair sheep: A general description," in *Hair sheep of western africa and the americas: A genetic Resource for the tropics*. Editors H. A. Fitzhugh and G. E. Bradford (Boulder, CO: Westview Press), 3. doi:10.1201/9780429049118

Bromley, C. M., Snowder, G. D., and Van Vleck, L. D. (2000). Genetic parameters among weight, prolificacy, and wool traits of Columbia, Polypay, Rambouillet, and Targhee sheep. *J. Animal Sci.* 78 (4), 846–858. doi:10.2527/2000.784846x

Burton, D. J., Ludden, P. A., Stobart, R. H., and Alexander, B. M. (2015). 50 years of the Wyoming ram test: How sheep have changed. *J. Animal Sci.* 93 (3), 1327–1331. doi:10.2527/jas.2014-8150

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., et al. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* 38 (7), 813–818. doi:10.1038/ng1810

Cockett, N. E., Shay, T. L., Beever, J. E., Nielsen, D., Albretsen, J., Georges, M., et al. (1999). Localization of the locus causing Spider Lamb Syndrome to the distal end of ovine Chromosome 6. *Mamm. Genome* 10 (1), 35–38. doi:10.1007/s003359900938

Davenport, K. M., Bickhart, D. M., Worley, K., Murali, S. C., Salavati, M., Clark, E. L., et al. (2022). An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *GigaScience* 11, giab096. doi:10.1093/gigascience/giab096

Davenport, K. M., Hiemke, C., McKay, S. D., Thorne, J. W., Lewis, R. M., Taylor, T., et al. (2020). Genetic structure and admixture in sheep from terminal breeds in the United States. *Anim. Genet.* 51 (2), 284–291. doi:10.1111/age.12905

Dong, S., Ying, S., Kojima, T., Shiraiwa, M., Kawada, A., Méchin, M. C., et al. (2008). Crucial roles of MZF1 and Sp1 in the transcriptional regulation of the peptidylarginine deiminase type I gene (PADI1) in human keratinocytes. *J. investigative dermatology* 128 (3), 549–557. doi:10.1038/sj.jid.5701048

Duguma, G., Cloete, S. W. P., Schoeman, S. J., and Jordaan, G. F. (2002). Genetic parameters of testicular measurements in Merino rams and the influence of scrotal circumference on total flock fertility. *South Afr. J. animal Sci.* 32 (2), 76–82. doi:10.4314/sajas.v32i2.3748

Ebrahimi, F., Gholizadeh, M., Rahimi-Mianji, G., and Farhadi, A. (2017). Detection of QTL for greasy fleece weight in sheep using a 50 K single nucleotide polymorphism chip. *Trop. animal health Prod.* 49 (8), 1657–1662. doi:10.1007/s11250-017-1373-x

Gong, H., Zhou, H., Forrest, R. H., Li, S., Wang, J., Dyer, J. M., et al. (2016). Wool keratin-associated protein genes in sheep—A review. *Genes* 7 (6), 24. doi:10.3390/genes7060024

Gou, Y., Vemaraju, S., Sweet, E. M., Kwon, H. J., and Riley, B. B. (2018). sox2 and sox3 Play unique roles in development of hair cells and neurons in the zebrafish inner ear. *Dev. Biol.* 435 (1), 73–83. doi:10.1016/j.ydbio.2018.01.010

Hanford, K. J., Van Vleck, L. D., and Snowder, G. D. (2005). Estimates of genetic parameters and genetic change for reproduction, weight, and wool characteristics of Rambouillet sheep. *Small Ruminant Res.* 57 (2-3), 175–186. doi:10.1016/j.smallrumres.2004.07.003

Hilcenko, C., Simpson, P. J., Finch, A. J., Bowler, F. R., Churcher, M. J., Jin, L., et al. (2013). Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. *J. Am. Soc. Hematol.* 121 (6), 1028–1038. doi:10.1182/blood-2012-10-461491

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical J.* 50 (3), 346–363. doi:10.1002/bimj.200810425

International Wool Textile Organization (IWTO) (2013). *IWTO-47–2013 measurement of the mean and distri- bution of fibre diameter of wool using an Optical Fibre Diameter Analyser (OFDA)*. Brussels, Belgium: IWTO.

Itenge, T. O. (2021). Application of PCR technique to detect polymorphism of the KRTAP1. 1 gene in three sheep breeds-A review. *Anal. Chemistry-Advancement, Perspect. Appl.* doi:10.5772/intechopen.96941

Ivanova, T., Stoikova-Grigorova, R., Bozhilova-Sakova, M., Ignatova, M., Dimitrova, I., and Koutev, V. (2021). Phenotypic and genetic characteristics of fecundity in sheep. A review. *Bulg. J. Agric. Sci.* 27 (5), 1002–1008.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42 (4), 348–354. doi:10.1038/ng.548

Kassambara, A. (2020b). Ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0 Available at: https://CRAN.R-project.org/package=ggpubr.

Kassambara, A. (2020a). rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.6. 0.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., Van Der Lee, R., et al. (2018). Jaspar 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids Res.* 46 (D1), D260–D266. doi:10.1093/nar/gkx1126

Khan, M. J., Abbas, A., Ayaz, M., Naeem, M., Akhter, M. S., and Soomro, M. H. (2012). Factors affecting wool quality and quantity in sheep. *Afr. J. Biotechnol.* 11 (73), 13761–13766. doi:10.5897/AJBX11.064

Krivoruchko, A., Yatsyk, O., and Kanibolockaya, A. (2022). New candidate genes of high productivity in North-Caucasian sheep using genome-wide association study (GWAS). *Anim. Gene* 23, 200119. doi:10.1016/j.angen.2021.200119

Kuramoto, T., Morimura, K., Nomoto, T., Namiki, C., Hamada, S., Fukushima, S., et al. (2005). Sparse and wavy hair: A New model for hypoplasia of hair follicle and mammary glands on rat chromosome 17. *J. Hered.* 96 (4), 339–345. doi:10.1093/jhered/esi053

Kustatscher, G., Grabowski, P., Schrader, T. A., Passmore, J. B., Schrader, M., and Rappsilber, J. (2019). Co-regulation map of the human proteome enables identification of protein functions. *Nat. Biotechnol.* 37 (11), 1361–1371. doi:10.1038/s41587-019-0298-5

Liu, G., Liu, R., Li, Q., Tang, X., Yu, M., Li, X., et al. (2013). Identification of microRNAs in wool follicles during anagen, catagen, and telogen phases in Tibetan sheep. *PloS one* 8 (10), e77801. doi:10.1371/journal.pone.0077801

Liver Marketing Information Center (LMIC) (2016). Sheep cost of production study. Available at: http://lmic.info/page/cost-sheep-produc- tion-budget-sponsored-american-sheep-industry. [accessed October 10, 2022].

Luo, W., and Yang, J. (2022). Schizophrenia predisposition gene Unc-51-like kinase 4 for the improvement of cerebral ischemia/reperfusion injury. *Mol. Biol. Rep.* 49, 2933–2943. doi:10.1007/s11033-021-07108-z

Lupton, C. J., Huston, J. E., Craddock, B. F., Pfeiffer, F. A., and Polk, W. L. (2007). Comparison of three systems for concurrent production of lamb meat and wool. *Small Ruminant Res.* 72 (2-3), 133–140. doi:10.1016/j.smallrumres.2006.10.002

Lupton, C. J., Waldron, D. F., and Pfeiffer, F. A. (2002). Interrelationships of traits measured on fine-wool rams during a central performance test. *Sheep Goat Res. J.* 18, 1–7.

Ma, G. W., Chu, Y. K., Zhang, W. J., Qin, F. Y., Xu, S. S., Yang, H., et al. (2017). Polymorphisms of FST gene and their association with wool quality traits in Chinese Merino sheep. *PLoS One* 12 (4), e0174868. doi:10.1371/journal.pone.0174868

Mahajan, V., Das, A. K., Taggar, R. K., Kumar, D., and Khan, N. (2019). Keratin-associated protein (kap) 8 gene polymorphism and its association with wool traits in rambouillet sheep. *Int. J. Life Sci. Appl. Sci.* 1 (1), 50.

Mahajan, V., Das, A. K., Taggar, R. K., Kumar, D., Khan, N., Sharma, R., et al. (2018). Effect of non-genetic factors on some wool traits in Rambouillet sheep. *Int. J. Curr. Microbiol. Appl. Sci.* 7, 3958–3965.

Mahajan, V., Das, A. K., Taggar, R. K., Kumar, D., and Sharma, R. (2017a). Association of polymorphic variants of KAP 1.3 gene with wool traits in Rambouillet sheep. *Indian J. Animal Sci.* 87 (10), 1237–1242.

Mahajan, V., Das, A. K., Taggar, R. K., Kumar, D., and Sharma, R. (2017b). Polymorphism of keratin-associated protein (KAP) 7 gene and its association with wool traits in Rambouillet sheep. *Indian J. Animal Sci.* 88, 206–209.

Medrado, B. D., Pedrosa, V. B., and Pinto, L. F. B. (2021). Meta-analysis of genetic parameters for economic traits in sheep. *Livest. Sci.* 247, 104477. doi:10.1016/j.livsci.2021.104477

Mohammadi, H., Rafat, S. A., Moradi Shahrbabak, H., Shodja, J., and Moradi, M. H. (2020). Genome-wide association study and gene ontology for growth and wool characteristics in Zandi sheep. *J. Livest. Sci. Technol.* 8 (2), 45–55. doi:10.22103/JLST.2020.15795.1317

Mu, F., Rong, E., Jing, Y., Yang, H., Ma, G., Yan, X., et al. (2017). Structural characterization and association of ovine Dickkopf-1 gene with wool production and quality traits in Chinese Merino. *Genes* 8 (12), 400. doi:10.3390/genes8120400

Murphy, T. W., Stewart, W. C., Notter, D. R., Mousel, M. R., Lewis, G. S., and Taylor, J. B. (2019). Evaluation of rambouillet, polypay, and romanov–white Dorper× rambouillet ewes mated to terminal sires in an extensive rangeland production system: Body weight and wool characteristics. *J. Animal Sci.* 97 (4), 1568–1577. doi:10.1093/jas/skz070

Naidoo, P., Olivier, J. J., Cloete, S. W. P., and Morris, J. (2016). Does selecting for finer wool result in higher incidence of creeping belly in the South African Dohne Merino sheep breed? *Agriprobe* 13 (2), 49–53.

NCBI Resource Coordinators (2016). Database resources of the national center for biotechnology information. *Nucleic acids Res.* 44 (D1), D7–D19. doi:10.1093/nar/gkv1290

Pedersen, T. (2020). patchwork: The composer of plots. R package version 1.1.1 Available at: https://CRAN.R-project.org/package=patchwork.

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

Qi, J. C., Jiang, Q. Q., Ma, L., Yuan, S. L., Sun, W., Yu, L. S., et al. (2022). Sox10 gene is required for the survival of saccular and utricular hair cells in a porcine model. *Mol. Neurobiol.* 59 (6), 3323–3335. doi:10.1007/s12035-021-02691-5

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rong, E. G., Yang, H., Zhang, Z. W., Wang, Z. P., Yan, X. H., Li, H., et al. (2015). Association of methionine synthase gene polymorphisms with wool production and quality traits in Chinese Merino population. *J. animal Sci.* 93 (10), 4601–4609. doi:10.2527/jas.2015-8963

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd Edn Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell stem Cell* 12 (1), 15–30. doi:10.1016/j.stem.2012.12.007

Selvaggi, M., Laudadio, V., Dario, C., and Tufarelli, V. (2015). β-Lactoglobulin gene polymorphisms in sheep and effects on milk production traits: A review. *Adv. Anim. Vet. Sci.* 3 (9), 478–484. doi:10.14737/journal.aavs/2015/3.9.478.484

Shelton, M., Miller, J. C., Magee, W. T., and Hardy, W. T. (1954). A summary of four years work in ram progeny and performance testing. *J. animal Sci.* 13 (1), 215–228. doi:10.2527/jas1954.131215x

Shorter, K., Farjo, N. P., Picksley, S. M., and Randall1, V. A. (2008). Human hair follicles contain two forms of ATP-sensitive potassium channels, only one of which is sensitive to minoxidil. *FASEB J.* 22 (6), 1725–1736. doi:10.1096/fj.07-099424

Singh, V. P., Taggar, R. K., Chakraborty, D., Pratap, B., Singh, P. K., Singh, S., et al. (2022). KRT 1.2 gene polymorphism & its association with wool traits in Rambouillet sheep. *Pharma Innovation J. SP-* 11 (6), 2619–2621.

SNP (2022), Snp & variation suite ™, Version 8.9.1, [Software]. Bozeman, MT: Golden Helix, Inc. Available at: http://www.goldenhelix.com.

Snyman, M. A., and Olivier, W. J. (2002). Correlations of subjectively assessed fleece and conformation traits with production and reproduction in Afrino sheep. *South Afr. J. Animal Sci.* 32 (2), 88–96. doi:10.4314/sajas.v32i2.3750

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., and Huerta-Cepas, J. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research* 43 (D1), D447–D452. doi:10.1093/nar/gku1003

Vesely, J. A., Peters, H. F., Slen, S. B., and Robison, O. W. (1970). Heritabilities and genetic correlations in growth and wool traits of Rambouillet and Romnelet sheep. *J. animal Sci.* 30 (2), 174–181. doi:10.2527/jas1970.302174x

Wagner, N., Panelos, J., Massi, D., and Wagner, K. D. (2008). The Wilms' tumor suppressor WT1 is associated with melanoma proliferation. *Pflügers Archiv-European J. Physiology* 455 (5), 839–847. doi:10.1007/s00424-007-0340-1

Wang, Z., Zhang, H., Yang, H., Wang, S., Rong, E., Pei, W., et al. (2014). Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PloS one* 9 (9), e107101. doi:10.1371/journal.pone.0107101

Wei, T., and Simko, V. (2021). R package 'corrplot': Visualization of a correlation matrix. Version 0.92.

Westaway, D., Zuliani, V., Cooper, C. M., Da Costa, M., Neuman, S., Jenny, A. L., et al. (1994). Homozygosity for prion protein alleles encoding glutamine-171 renders sheep susceptible to natural scrapie. *Genes & Dev.* 8 (8), 959–969. doi:10.1101/gad.8.8.959

Wilson, D. E., and Morrical, D. G. (1991). The national sheep improvement program: A review. *J. animal Sci.* 69 (9), 3872–3881. doi:10.2527/1991.6993872x

Yang, X., Chen, Y., Liu, X., Liu, Q., Pi, X., Liu, Y., et al. (2016). De novo characterization of velvet skin transcriptome at the antlers tips of red deer (*Cervus elaphus*) and analysis of growth factors and their receptors related to regeneration. *Pak. J. Zoology* 48 (1).

Yin, L. (2022). CMplot: Circle manhattan plot. R package version 4.1.0 Available at: https://CRAN.R-project.org/package=CMplot.

Zhao, H., Guo, T., Lu, Z., Liu, J., Zhu, S., Qiao, G., et al. (2021a). Genome-wide association studies detects candidate genes for wool traits by re-sequencing in Chinese fine-wool sheep. *BMC genomics* 22, 127. doi:10.1186/s12864-021-07399-3

Zhao, B., Luo, H., Huang, X., Wei, C., Di, J., Tian, Y., et al. (2021b). Integration of a single-step genome-wide association study with a multi-tissue transcriptome analysis provides novel insights into the genetic basis of wool and weight traits in sheep. *Genet. Sel. Evol.* 53 (1), 56–14. doi:10.1186/s12711-021-00649-8

Zhao, H., Hu, R., Li, F., and Yue, X. (2021c). Two strongly linked blocks within the KIF16B gene significantly influence wool length and greasy yield in fine wool sheep (*Ovis aries*). *Electron. J. Biotechnol.* 53, 23–32. doi:10.1016/j.ejbt.2021.05.003

Zhao, H., Hu, R., Li, F., and Yue, X. (2021d). Five SNPs within the FGF5 gene significantly affect both wool traits and growth performance in fine-wool sheep (*Ovis aries*). *Front. Genet.* 12, 732097. doi:10.3389/fgene.2021.732097

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 April 2023
DOI 10.3389/fgene.2023.1092066

Check for updates

OPEN ACCESS

EDITED BY
Ibrar Muhammad Khan,
Fuyang Normal University, China

REVIEWED BY
Hongyu Liu,
Anhui Agricultural University, China
Samiullah Khan,
Guizhou University, China

*CORRESPONDENCE
Shudong Liu,
liushudong63@126.com

†These authors have contributed equally to this work

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

RECEIVED 07 November 2022
ACCEPTED 28 March 2023
PUBLISHED 11 April 2023

CITATION
Zhang C-l, Zhang J, Tuersuntuoheti M, Chang Q and Liu S (2023), Population structure, genetic diversity and prolificacy in pishan red sheep under an extreme desert environment.
Front. Genet. 14:1092066.
doi: 10.3389/fgene.2023.1092066

COPYRIGHT
© 2023 Zhang, Zhang, Tuersuntuoheti, Chang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Population structure, genetic diversity and prolificacy in pishan red sheep under an extreme desert environment

Cheng-long Zhang[1,2†], Jihu Zhang[1,2†], Mirenisa Tuersuntuoheti[1,2], Qianqian Chang[1,2] and Shudong Liu[1,2*]

[1]College of Animal Science and Technology, Tarim University, Alar, China, [2]Key Laboratory of Tarim Animal Husbandry Science and Technology, Xinjiang Production and Construction Corps, Alar, China

Extreme environmental conditions are a major challenge for livestock production. Changes in climate conditions, especially those that lead to extreme weather, can reduce livestock production. The screening of genes and molecular markers is of great significance to explore the genetic mechanism of sheep prolificacy traits in Taklimakan Desert environment. We selected healthy adult Pishan Red Sheep (PRS) and Qira Black Sheep (QR) which live in Taklimakan Desert environment, collected blood from jugular vein, extracted DNA, and prepared Illumina Ovine SNP50 chip. For PRS, linkage disequilibrium (LD) was calculated using the ovine SNP50 Beadchip and the effective population size (*Ne*) was estimated using SMC++. The genetic characteristics of PRS were analyzed by integrated haplotype score (iHS) and fixation index (F_{ST}). The result showed that r^2 of PRS was 0.233 ± 0.280 in the range of 0–10 Kb and decreased with increasing distances. SMC++ tested that the *Ne* of PRS remained at 236.99 in recent generations. 184 genes were screened out under iHS 1% threshold, and 1148 genes were screened out with F_{ST} under the 5% threshold, and 29 genes were obtained from the intersection of the two gene sets. In this study, the genetic characteristics of PRS and QR were compared by ovine genome chip, and the related excellent genes were searched, providing reference for the protection of sheep germplasm resources and molecular breeding in a desert environment.

KEYWORDS

desert environment, genomic selection, linkage disequilibrium, perennial estrus, litter size

1 Introduction

Sheep is one of the earliest domesticated animals in the world, and also one of the most successful animals domesticated by human beings in the Neolithic age. After long-term domestication and different environments, sheep have great changes in morphology, physiology and behavior. Pishan Red sheep (PRS) living in the Taklimakan Desert is characterized by perennial estrus and multiple fetuses after long-term selection by nature and people. In addition to its high-quality production traits, PRS is also a rare breed of sheep because its origin is on the southern edge of the Taklimakan Desert and north of the Karakoram Mountains (Lv et al., 2020). Pishan red sheep is a local sheep breed formed under the local cultural and geographical conditions. The origin and formation history of PRS is not

fully understood. Therefore, to dig the genetic structure and molecular markers of PRS can better protect PRS.

Linkage disequilibrium (LD) can improve the accuracy of genomic association analysis and predict marker regions (Liu et al., 2017). LD decay patterns also provide information about the evolutionary history of the population and can be used to estimate ancestral effective population size (*Ne*) (Tenesa et al., 2007). *Ne* and other genetic events can also influence the extent of LD in the population (Wang, 2005). Therefore, LD helps in understanding the selection patterns experienced by individual breeds. Currently, LD estimates have been reported in several studies for a variety of livestock species, such as cattle (Porto-Neto et al., 2014), pigs (Uimari and Tapio, 2011), horses (Corbin et al., 2010), chickens (Rao et al., 2008) and sheep (Kijas et al., 2014).

The selected regions were searched through different chromosomes to provide a molecular genetic basis for sheep protection. Compared with traditional selection methods, genomics can be evaluated early with higher accuracy (Hayes et al., 2013). Voight and Kudaravalli proposed an integrated haplotype score (iHS) test based on extended haplotype homozygosity (Voight et al., 2006). The incomparability of test statistics caused by differences in recombination rates between different chromosome segments was corrected by calculating the Extended Haplotype Homozygosity (EHH) statistics and genetic distance integration. Fixation index (F_{ST}) is used to measure the degree of population differentiation, indicating that there are obvious allele frequency differences between populations (Akey et al., 2002). Genetic drift and selection process can usually cause the genetic differentiation between populations. This method is suitable for selection signal detection of multiple populations. Now, it is necessary to selectively intervene in the breeding work of PRS through genomic selection technology to guarantee a better inheritance of the excellent production traits of PRS to future generations.

In order to analyze the genetic mechanism of prolific traits in PRS under desert environment, we selected prolific PRS and singleton pregnancy Qira Black Sheep (QR) under relevant survival background as research objects. Genetic basis of PRS was analyzed based on LD and *Ne* and the genetic mechanism of prolific traits in Taklimakan Desert was explored by using genomic selection method.

2 Materials and methods

2.1 Animal care

This work was conducted in accordance with the specifications of the Ethics Committee of Tarim University of Science and Technology (SYXK 2020-009).

2.2 Animal collection

33 PRS (polyembryony) and 40 QR (singleton pregnancy) were randomly selected from Pishan County and Cele County in Hetian region and all of them were healthy adult sheep with no genetic relationship. Blood samples were collected from the jugular vein and DNA was extracted with a DNA kit (Tiangen Biotech Co. Ltd., Beijing, China). The samples were sent to Beijing Compass Agritechnology Co., Ltd. to prepare the Illumina Ovine SNP50 Beadchip (The number of SNPs in this chip is 50 k). The validation samples were 130 PRS (first lambing) from Pishan Farm, with the same feeding conditions, 1.5–2.5 years old, no relationship.

2.3 Genotyping and data quality control

Genome Studio software was used to process the preliminary data results and obtain the VCF files. Plink (Purcell et al., 2007) software was used for quality control. Unqualified SNP sites were eliminated. The quality control criteria of this study were as follows: 1) individual detection rate >0.95, 2) SNP detection rate >0.95, and 3) Hardy-Weinberg equilibrium (HWE) with p values $\geq 10^{-6}$.

2.4 LD calculation method

LD is the basis of association analysis, and the analysis of LD between loci helps to understand the LD level of the PRS genome. Since r^2 is more capable of objectively reflecting the LD between different loci, it was adopted in this study as the LD measurement standard. The LD values ranged from 0 to 1. As the LD level increased with the value of r^2, the linkage degree increased. The calculation formula of r^2 is as follows (Hill, 1974):

$$r^2 = \frac{(P_{A1B1} - P_{A1} \cdot P_{B1})^2}{P_{A1} \cdot (1 - P_{A1}) \cdot P_{B1} \cdot (1 - P_{B1})}$$

Where PA1 and PB1 are the frequency of the first allele at the two marker loci and the haplotype frequency formed between alleles. The correlation coefficient (r^2 mean) of alleles was calculated to measure the level of linkage disequilibrium (LD) using PopLDdecay V3.41 (Zhang et al., 2018), and perl scripts were used to visualize the results.

2.5 Estimation of effective population size

The SMC++ (Terhorst et al., 2016) method was used to estimate *Ne*. The population size history and splitting time of the PRS can be predicted with SMC++. A new spline regularization scheme was adopted in this method, significantly reducing estimation errors. The conversion of each VCF file into an input file in SMC++ format was made using the vcf2smc script distributed by SMC++. All the simulations were performed under the initial condition of a mutation rate of 1.25×10^{-8}.

2.6 Genetic diversity and population structure

The genotypic data after quality control was subjected to Principal Component Analysis (PCA) using MingPCACluster (https://github.com/hewm2008/MingPCACluster). The VCF2Dis

v1.09 (https://github.com/BGI-shenzhen/VCF2Dis) was used to calculate the P distance matrix, and then the NJ-tree was constructed by ATGC:FastME (http://www.atgc-montpellier.fr/fastme) program. Genetic admixture calculations were performed using Admixture (Patterson et al., 2012).

2.7 Fixation index

F_{ST} is used to measure the degree of population differentiation and can reflect the level of species population differentiation. This method is suitable for selective signal detection of multiple populations and as follows:

$$F_{ST} = \frac{MSP - MSG}{MSP + (nc - 1)MSG}$$

Where, MSG is the mean square of error within the population, MSP is the mean square of error between the populations, and nc is the average sample size between the populations after correction. By using a sliding window with a window size of 50 Kb and a sliding step size of 25 Kb, the F_{ST} value of each sliding window SNP is calculated. Vcftools was used to calculate the F_{ST} value for each window, and then CMplot was used to plot Manhattan Yin et al. (2021).

2.8 Integrated haplotype score

An intrapopulation selective genomic sweep analysis was performed on all individuals using iHS. iHS is an alternative EHH statistic using a single marker loci haploid type. It is defined as the core in the site, the expansion of the ancestors in the core loci alleles in the haploid type, and new mutant alleles in the extension of the haploid type EHH statistics for the integral genetic distance. It is possible to calculate the ratio between the previously mentioned genetic metrics to select a signal detection statistic using this method, as expressed by Voight et al. (2006):

$$IHS = \frac{unIHS - mean(unIHS \mid ps)}{sd(unIHS \mid ps)}$$

uniHS is:

$$unIHS = In\left(\frac{IHH_A}{IHH_D}\right)$$

Where IHH (Integrated EHH) refers to integrating genetic distance with EHH; A is the ancestral allele, and D is the newly derived allele.

2.9 Enrichment analysis of candidate genes

The iHS results and F_{ST} results were selected for intersection analysis, with annotations with the sheep genome Ovis Oar_v4.0. Gene functional annotation was performed referencing the NCBI databases (http://www.ncbi.nlm.nih.gov/gene) and OMIM database (http://www.ncbi.nlm.nih.gov/omim). The g:Profiler (https://biit.cs.ut.ee/gprofiler/gost) was used for autosomal enrichment of candidate genes for GO and Reactome/KEGG pathway analysis.

2.10 PCR amplification of BMPR1B

The Fec^B locus of the BMPR1B gene in 130 PRS was amplified by PCR according to the primers shown in Supplementary Material. After the PCR products were detected by 1.5% agarose gel electrophoresis, all qualified PCR products were sent to Beijing Compass Agritechnology Co., Ltd. for DNA sequencing and genotype identification.

3 Results

3.1 Descriptive statistics

Genotypic quality control was conducted on the SNPs of the 33 PRS and 40 QR used in the experiment. After the unqualified SNPs were removed, there were 49,219 informative SNPs in the PRS and QR population.

3.2 The extent of genome-wide LD and effective population size of PRS

When LD was calculated, the distance between markers was set in the 0–1 Mb (0–10, 10–25, 25–50, 50–100, 100–500 Kb, 0.5–1 Mb) autosomal range (Table 1). The average r^2 decreased with increasing physical distance (Figure 1A). When the distance between markers was 10–25 Kb, the average r^2 value was 0.151 ± 0.200. Compared with the average r^2 value between SNPs within 0–10 Kb (0.233 ± 0.280), the difference was 0.082, much larger than the average r^2 value between other adjacent distance regions. When the distance increased to 50–100 Kb, the r^2 value was lower than 0.1, indicating that the LD was weak.

Based on LD, six generations *Ne* of PRS were estimated (Figure 1B). The PRS remained relatively stable at about 631.97 during the 2,500–2,000 generations. It rose to 1,119.99 with 1,500 generations, 1,328.81 with 1,000 generations, and fell to 898.12 with 500 generations. It declined even faster until recent generations stood at 236.99. During the first 200-100 years, PRS populations remained relatively stable. It was only in the last 10 years that the population increased slightly due to the introduction of measures to protect the endemic species.

3.3 Genetic diversity and population structure

PCA analysis can separate PRS from QR, and part of PRS extended outward (Figure 2A). In Figure 2A, k-means clustering was carried out according to the predefined subsets. When k = 2, the distinction was obvious. Neighbor-Joining (N-J) Tree showed that PRS and QR were divided into two varieties (Figure 2B), which was consistent with PCA analysis. Admixture analysis showed that PRS and QR had similar genetic backgrounds (Figure 2C).

3.4 Selective gene sweep

The F_{ST} values of PRS and QR were calculated, and the F_{ST} values were arranged in descending order. The first 5% was regarded as the significant window (Figure 3A), and a total of 1148 genes were obtained.

TABLE 1 LD statistical analysis between different distances (0–1 Mb).

Distance	Average r^2	Number of SNP pairs	Proportion of $r^2 > 0.2$	Proportion of $r^2 > 0.3$
0-10 Kb	0.233 ± 0.280	8561	0.364	0.288
10-25 Kb	0.151 ± 0.200	10,948	0.233	0.150
25-50 Kb	0.120 ± 0.158	18,123	0.177	0.099
50-100 Kb	0.096 ± 0.115	36,123	0.122	0.050
100-500 Kb	0.078 ± 0.080	185,353	0.075	0.019
0.5-1 Mb	0.074 ± 0.073	368,857	0.062	0.014

r^2: denotes the extent of LD.

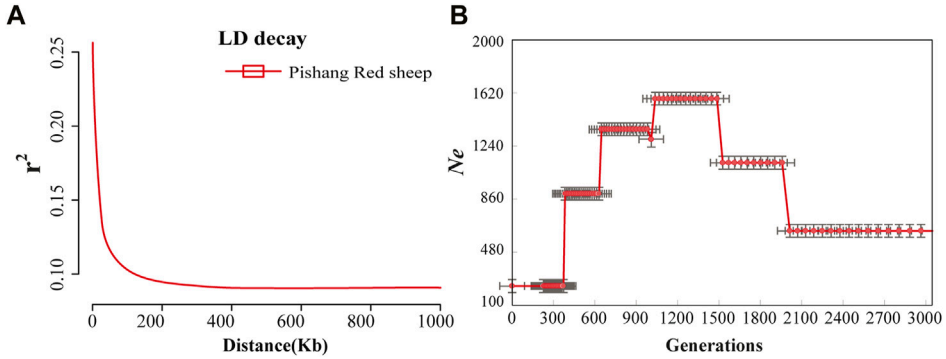


FIGURE 1
LD decay and *Ne* in PRS. **(A)**: Distribution of average r^2 values for PRS with respect to physical distance (kb) in 1,000 kb windows. **(B)**: Genome-wide LD (r^2) between SNPs was used to estimate the *Ne* of the PRS.

Under the 1% threshold, 184 candidate genes were screened out with iHS (Figure 3B). A total of 29 genes were obtained from the intersection of the two gene sets (Figure 3C), scuh as *BMPR1B*, *3BHSD*, *STPG2*, *ATRN*, *GANS*, etc (*Tabel A2. xlsx*).

3.5 Functional annotation of genes

A total of 184 genes were screened by iHS and GO and Reactome pathway analysis were performed on the 184 genes. The enrichment results of the Reactome pathway of the candidate genes were shown in Figure 4. Some genes were related to animal reproduction (Table 2). Genes *ABCA4*, *ARL4C*, and *PARP14* can affect the function of ion channels. Genes *SOX2*, *DAB1*, and *COL5A2* regulate cell differentiation. 29 genes were obtained by the intersection of the results of F_{ST} and IHS, and after pathway enrichment, these genes were found to be related to GnRH signalingpathway, Ovarian steroidogenesis and Estrogen signaling pathway (Figure 5), indicateing that these genes may control PRS prolific trait.

3.6 PCR product sequencing and genotyping

After 1.5% agarose gel electrophoresis, the PCR products were found to conform to the expected size, indicating that the target fragment was successfully amplified (Figure 6A). The sequencing results are shown in Figure 6B. SNP locus detection revealed that BMPR1B gene G.431965A > G, and three genotypes, B+, ++, and BB, were detected (Table 3). The number of lambs of the BB genotype was significantly higher than that of the B+ and ++ genotypes ($p < 0.05$). The number of lambs of the B+ genotype was significantly higher than that of the ++ genotype ($p < 0.05$). It was shown in the χ^2 test that the PRS reached Hardy-Weinberg equilibrium at the Fec^B locus.

4 Discussion

The Pishan Red sheep is a newly discovered local sheep group known for its stress resistance, perennial estrus, and high fecundity. Understanding the breed characteristics and evolutionary history of PRS will help to further strengthen the conservation and utilization of its genetic resources. PCA, N-J tree and Admixture suggest that PRS and QR could be subdivided into two genetic clusters, and have similar ancestral components.

4.1 LD and *Ne* population size

In PRS, LD was only moderate at 0–10 Kb and rapidly decreased to 0.120 ± 0.158 at 25–50 Kb. The r^2 we calculated for PRS was close

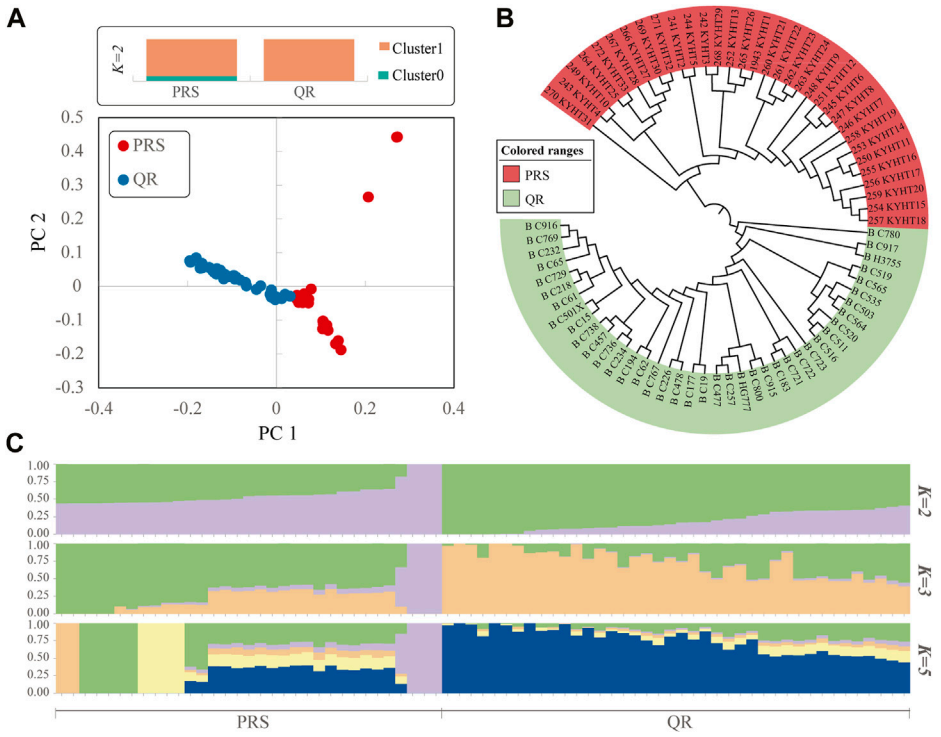


FIGURE 2
Analysis of the population structure in PRS and QR. **(A)**, Principal component analysis, k-means clustering was carried out according to the predefined subsets. **(B)**, Neighbor-joining tree. **(C)**, Analysis of Admixture with the assumed number (K = 2, 3 and 5).

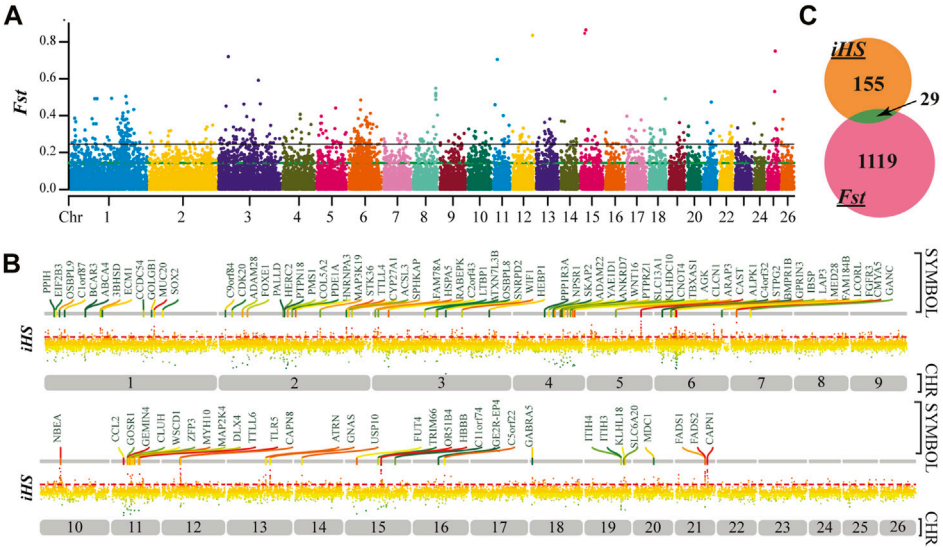


FIGURE 3
Selective gene sweep. **(A)**, F_{ST} detection results of PRS and QR. The black line represents the 1% threshold line and the green line represents the 5% threshold line. **(B)**, The 100 high-rank genes under the iHS 1% threshold. **(C)**, The intersection of F_{ST} and IHS results.

to that of other sheep breeds and species. In Iranian Zandi sheep, the average r^2 for the pairwise space of 0–10 Kb was 0.26 (Ghoreishifar et al., 2019). In the Barbaresca sheep, the average r^2 for the intermarker distance of 0.5–1.0 Mb was 0.12 (Mastrangelo et al., 2017). However, in Border Leicester and Poll Dorset, the average r^2 for the intermarker distance of 0–10 Kb was 0.34 and 0.33, respectively (Al-Mamun et al., 2015). The average r^2 for the intermarker distances of 0–10 Kb and 10–25 Kb was 0.43 and

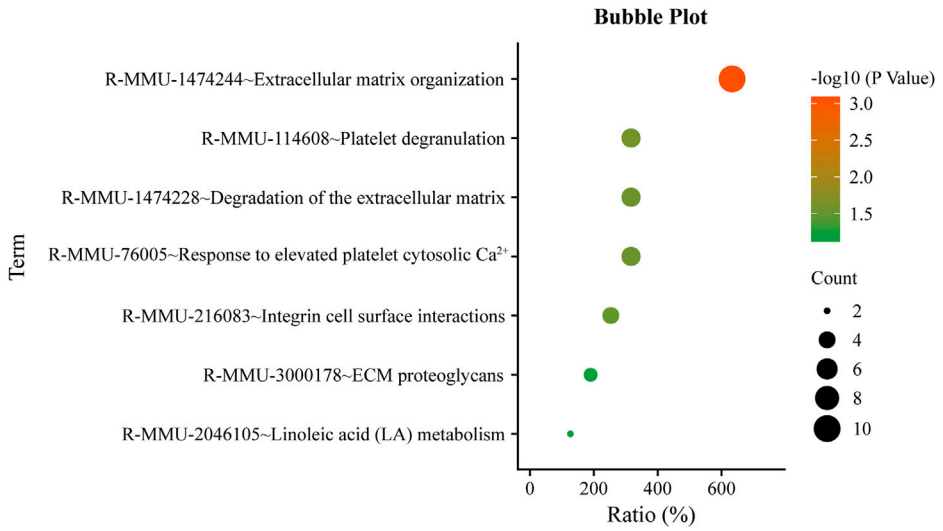


FIGURE 4
The Reactome pathways of 184 candidate genes screened out with iHS.

0.26, respectively, in Vrindavani cattle (Singh et al., 2021). In Gir cattle, the average r^2 value of 0.5–1 Mb was 0.032 (Ospina et al., 2019). Thus, the variation in the reported r^2 in different breeds and species suggests that LD are highly specific in sheep breeds. The PRS is located at the edge of the Taklimakan Desert. Its unique geographical environment may be the main reason for its high genetic diversity.

Mean r^2 values varied on chromosomes (from 0.182 ± 0.269 in OAR25 to 0.285 ± 0.367 in OAR19, distance <10 Kb), consistent with previous reports in sheep (Liu et al., 2017), beef cattle (Edea et al., 2014), and dairy cows (Qanbari et al., 2009). This phenomenon may be due to differences in recombination rates in different chromosomes, natural or artificial selection, and genetic drift (Mastrangelo et al., 2017). In the Vrindavani cattle, the highest mean r^2 value of chromosome 28 was 0.643, and the lowest r^2 value of chromosome 18 was 0.172 (0–10 Kb) (Singh et al., 2021). Moreover, the variation in r^2 estimated for different chromosomes was higher in short SNP pair distances, which is in line with the results reported for Chinese Merino sheep (Xinjiang type) (Liu et al., 2017). As seen from the attenuation of LD of 26 chromosomes of PRS, the LD of each chromosome was weak. The r^2 values were higher where the distance between the marked sites was close, but there were also higher r^2 values between the two sites, indicating a certain pattern of LD between the distant sites. The weak LD levels of OAR12, OAR18, OAR20, and OAR25 indicated that the degree of purification of these chromosomes was not high, leading to higher genetic diversity than other chromosomes. The chromosomes with higher r^2 values were OAR19, OAR15, OAR13, and OAR14, indicating that these four chromosomes may be more strongly selected than the other chromosomes.

The *Ne* of 236.99 of PRS 50 generations ago was similar to that of Chinese Merino sheep (Liu et al., 2017). We observed that *Ne* decreased more strongly from about 380 generations ago, consistent with the results of Chinese Merino sheep (Liu et al., 2017). The low level of r^2, even at relatively short distances, showed that the *Ne* in PRS was large in recent past generations compared with other species. For example, the r^2 for SNP pairs within 0.9–1.0 Mb and the *Ne* in recent generations in Duroc pigs were reported to be 0.2 and 75, respectively (Grossi et al., 2017). However, given the sharp drop in *Ne* in recent generations, we should be careful to maintain *Ne* larger than 100 individuals.

4.2 Adaptive mechanisms of the desert environment

The environmental adaptability differs between the Taklimakan Desert sheep breeds and those of other areas. PRS can adapt to extreme conditions, such as high salinity, drought, and ultraviolet rays. Under the iHS 1% threshold, 184 genes were screened. It was revealed that the genetic evidence and physiological mechanism of PRS adaptation to the desert environment. In terms of linoleic acid (LA) metabolism (R-MMU-2046105), *FADS1* and *FADS2* control polyunsaturated fatty acids (Hanson and Korotkova, 2002; Lawrence and Lawrence, 2004; Das, 2006; Calder, 2015), which accumulate fat to cope with extreme weather. The activation of phospholipase C enzymes results in the generation of second messengers of the phosphatidylinositol pathway in terms of adaptability. The events resulting from this pathway increase intracellular calcium and protein kinase C (PKC) activation. Phospholipase C cleaves the phosphodiester bond in PIP2 to form 1,2 diacylglycerol (DAG) and 1,4,5-inositol trisphosphate (IP3). IP3 opens Ca^{2+} channels in the platelet-dense tubular system, raising intracellular Ca^{2+} levels (R-MMU-76005). In terms of immunity, *HERC2* and *USP10* can promote the activation of the ATR-CHK1 pathway, triggering cell cycle checkpoints (Yuan et al., 2014). *PAALD* is involved in the process of phagocytosis (Sun et al., 2017). *KIF2A*, *TRIM66*, *BCL11B*, and *CLEC14A* play essential roles in repairing damaged cell DNA (Chen et al., 2019; Seira et al., 2019; Valisno et al., 2021; Su

TABLE 2 Candidate genes related to reproduction.

Gene symbol	NCBI gene ID	Go term name	Go term ID	OAR	Coordinates (bp)
PAX5	101108719	developmental process involved in reproduction	GO:0003006	2	51663045–51849640
		reproductive process	GO:0022414		
		sexual reproduction	GO:0019953		
HERC2	101102534	developmental process involved in reproduction	GO:0003006	2	113570525–113814058
		reproductive process	GO:0022414		
		Reproduction	GO:0000003		
		multicellular organism reproduction	GO:0032504		
		multi-organism reproductive process	GO:0044703		
HSPA5	780447	reproductive structure development	GO:0048608	3	10783159–10787445
		developmental process involved in reproduction	GO:0003006		
		reproductive process	GO:0022414		
		multicellular organismal reproductive process	GO:0048609		
INSR	443431	reproductive structure development	GO:0048608	5	13788125–13936266
		developmental process involved in reproduction	GO:0003006		
CAST	443364	reproductive process	GO:0022414	5	93695915–93785491
		sexual reproduction	GO:0019953		
BMPR1B	443454	reproductive structure development	GO:0048608	6	30030664–30482585
		developmental process involved in reproduction	GO:0003006		
		reproductive process	GO:0022414		
		cellular process involved in reproduction in multicellular organism	GO:0022412		
		Reproduction	GO:0000003		
		reproductive system development	GO:0061458		
		multicellular organism reproduction	GO:0032504		
		multi-organism reproductive process	GO:0044703		
GABRB1	101122358	Reproduction	GO:0000003	6	66336302–66776044
		multicellular organism reproduction	GO:0032504		
HAS2	101110341	multicellular organism reproduction	GO:0032504	9	31108300–31140164
TLR5	554256	reproductive structure development	GO:0048608	12	25835725–25861758
		reproductive system development	GO:0061458		
C11ORF74	101107893	developmental process involved in reproduction	GO:0003006	15	65840095–65907755
		reproductive process	GO:0022414		
		Reproduction	GO:0000003		
		multi-organism reproductive process	GO:0044703		
FGF10	443074	reproductive structure development	GO:0048608	16	30604216–30700637
		developmental process involved in reproduction	GO:0003006		
		reproductive process	GO:0022414		
		reproduction	GO:0000003		

(Continued on following page)

10.3389/fgene.2023.1092066

TABLE 2 (*Continued*) Candidate genes related to reproduction.

Gene symbol	NCBI gene ID	Go term name	Go term ID	OAR	Coordinates (bp)
ABHD2	101103014	developmental process involved in reproduction	GO:0003006	18	19910266–20020987
		reproductive process	GO:0022414		
		cellular process involved in reproduction in multicellular organism	GO:0022412		
		Reproduction	GO:0000003		
		sexual reproduction	GO:0019953		
		multicellular organism reproduction	GO:0048609		
		multicellular organismal reproductive process	GO:0044703		
		multi-organism reproductive process			
ZFP42	101110027	reproductive structure development	GO:0048608	26	16740249–16747149
		developmental process involved in reproduction	GO:0003006		
		reproductive process	GO:0022414		

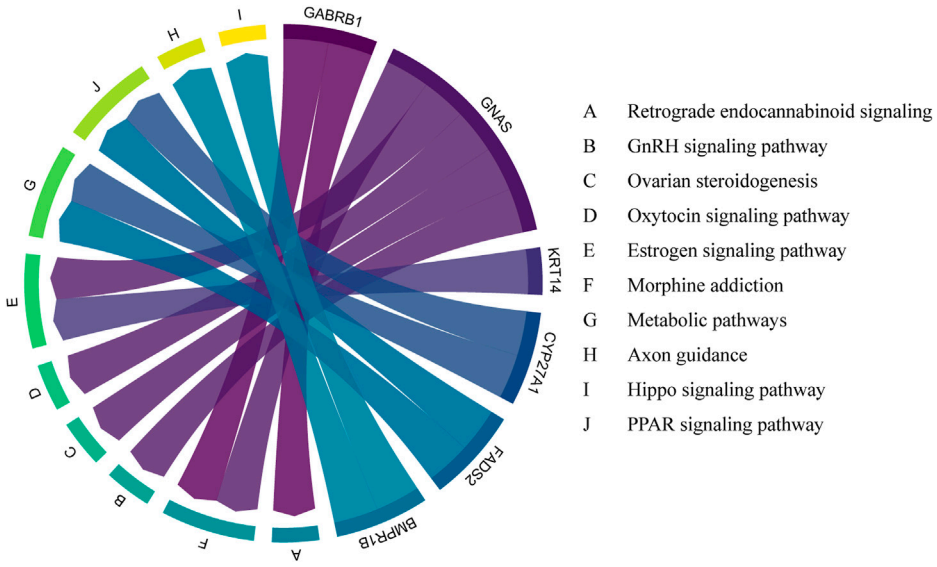


FIGURE 5
Enrichment pathways of 29 genes obtained by intersection of F_{ST} and IHS and their relationship.

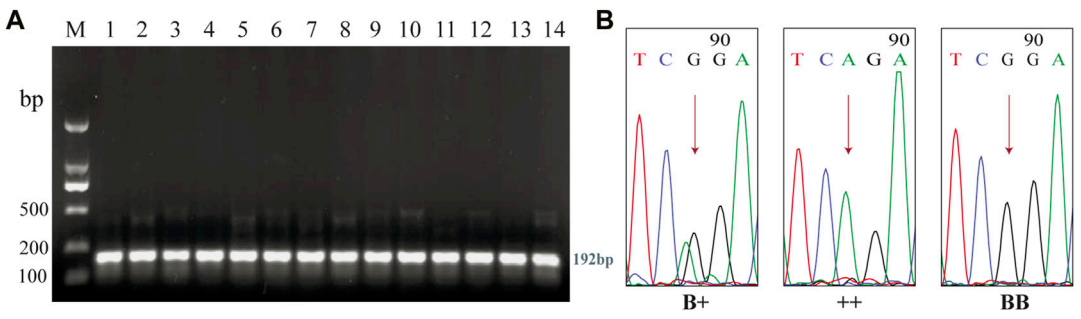


FIGURE 6
PCR amplification and sequencing of BMRPR1B. **(A)**, Detection of PCR amplification products of BMPR1B gene in PRS. M, DL2000DNA Marker. 1-14, PCR amplification product of *BMPR1B* gene. **(B)**, Sequencing results of the B+, ++, and BB genotypes.

TABLE 3 The genotype frequency and litter size of different BMPR1B genotypes of Pishan Red sheep.

Genotype	Number of samples	Genotype frequency	Allele frequency		Average number of lambs	χ^2
			B	+		
B+	64	0.492	0.5077	0.4923	1.453 ± 0.063[a]	1.1151
BB	34	0.262			1.735 ± 0.077[b]	
++	32	0.246			1.125 ± 0.059[c]	

Note: Different letters of shoulder labels of data in the same column indicate significant difference ($p < 0.05$).

et al., 2021). *GPRIN3* and *HERC3* control cell senescence and apoptosis (Chen et al., 2018; Ding et al., 2021). In terms of growth and development, *ZFP42* can control the differentiation of embryonic stem cells (Scotland et al., 2009). *EVC* can promote chondrogenesis (Lamuedra et al., 2022). *GABRB1* correlates with hypothalamic volume and regulates intelligence (Zhu et al., 2014).

4.3 Genetic mechanisms of perennial estrus and reproduction

The estrus cycle refers to the time between the previous and next ovulation. The ovary undergoes follicular growth, maturation, ovulation, luteal formation, and degeneration during the estrus cycle. The vast majority of sheep are singleton and seasonally estrus, leading to the failure of a balanced supply of lamb meat in the four seasons, seriously restricting the production efficiency of the meat sheep industry. PRS live in desert environments, and after natural and artificial selection, it forms the characteristics of perennial estrus and the early onset of puberty. *TLR5* can effectively alleviate the stimulation caused by radiation (Brackett et al., 2021), and *SOX10* independently regulates the expression of *IRF1* in melanoma through the JAK-STAT signaling pathway (Yokoyama et al., 2021). *ATP6V0A* can affect fetal brain development (Aoto et al., 2021). *AUH* can cause the early onset of puberty (Bizjak et al., 2020). During the luteal phase of the estrus cycle, IFNE is highly expressed in the uterus and has a protective effect against uterine infection (Fischer et al., 2018). As a member of the BMP system, *BMPR1B* plays a significant role in the sheep ovary. The BMP system can control the proliferation and differentiation of ovarian granulosa cells and the development of oocytes, among which *BMPR1B* plays an essential role in the regulation of ovarian function. The SNP of *BMPR1B* C.746 A > G, the 249th amino acid change, partially inactivates *BMPR1B* protein. This change affects the reaction of the ligands *GDF5* and *BMP4* recognized by *BMPR1B* to steroid production, making follicles mature earlier and increasing the ovine ovulation number (Kumar et al., 2020). *BMPR1B* is the primary gene affecting the trait of lambing abundance in sheep (Mulsant et al., 2001; Al-Samerria et al., 2015).

Fibroblast growth factor (FGF) is a large family of paracrine cells that can regulate follicular development and oocyte maturation. *FGF10* can interact with *BMP15* to increase cumulus cell diffusion and improve glucose utilization (Caixeta et al., 2013). *FGF10* was expressed in oocytes and membrane cells in cattle follicles, acting on granulosa cells to inhibit steroid production. *FGF10* regulates the cumulus-oocyte complex, improving expansion and development ability (Zhang et al., 2010; Zhang et al., 2020). *FGF10* can also reduce the proportion of apoptotic oocytes and increase the number of cells developing to the blastocyst stage (Pinto et al., 2014). In embryonic development, *FGF10* can activate the MAPK pathway, increase the phosphorylation level of MAPK, and mediate the migration of sheep trophoblast cells (Yang et al., 2011). *FGF10* can reduce the level of FSHR mRNA in granulosa cells *in vitro*. *FGF10* can also inhibit the secretion of estrogen. The injection of *FGF10* at the initial stage of follicular selection can inhibit follicular growth, which may be achieved by reducing FSHR mRNA levels and thus inhibiting estrogen expression. At the same time, the *FGF10* mRNA concentration in the ovaries of healthy and growing bovine follicles was higher than that in atresia follicles (Buratini et al., 2007), which indicated that *FGF10* played an inhibitory role in early follicular development and promoted follicular maturation in late follicular development. *FGF10* can also regulate the expression of genes *CD9*, *CD81*, *DNMT1*, and *DNMT3B* to improve embryo quality (Pan et al., 2019).

Hyaluronic acid (HA), also called hyaluronate or hyaluronan, is an essential extracellular matrix component widely available in various mammalian tissues (Rodgers and Irving-Rodgers, 2010). HA is synthesized by three HA synthase families (*HAS1*, *HAS2*, and *HAS3*). *HAS2* is so critical to development that the HAS2-HA system is considered an essential HAS-HA system. *HAS2* is responsible for the rapid hyaluronic acid synthesis in cumulus-oocyte complexes and granulosa cells. In the dominant follicles of mammals, most of the hyaluronic acid is secreted by cumulus cells and is also present in the follicular fluid. *HAS2* plays a vital physiological role in oocyte maturation, ovulation, *in vivo* fertilization, and early embryonic development (Irving-Rodgers and Rodgers, 2006). *C11orf74* enhances sperm motility and improves fertilization (Majkowski et al., 2018). Progesterone can upregulate *ABHD2* to activate the camp-PKA signaling pathway (Jiang et al., 2021). These genes were strongly selected, suggesting that perennial estrus may result from a combination of these genes.

The number of lambs is a complex character affected by many factors, among which heredity is the main factor. The Fec^B gene is the most studied gene and can significantly affect the ovulation number and lambing number of sheep breeds. The B+ genotype exists among multiple sheep breeds in Xinjiang, China, including Duolang and Hotan sheep. The number of lambs of the B+ genotype was significantly higher than that of the ++ genotype. We detected Fec^B in 130 leather-red sheep and found three genotypes: BB, B+, and ++. The average number of lambs born in the BB genotype was the highest, followed by the B+ and ++ genotypes, indicating that the Fec^B gene is an effective gene for improving the fecality of PRS.

5 Conclusion

In the extreme environment of the desert, PRS has formed the characteristics of perennial estrus, multiple pregnancies, and adequate resistance to stress. The LD decay rate of the PRS was faster, and LD and the *Ne* were at a relatively low level, suggesting that we should take reasonable protection measures to increase the PRS population. Genomic selection signal analysis and population validation showed that Fec^B could be used as a molecular breeding marker for multiple fetal lines of PRS in desert environments.

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: Our data has been uploaded to this website https://figshare.com/articles/dataset/e_vcf/21700937.

Author contributions

C-lZ and JZ led the bioinformatic and statistical analyses of data and helped to draft the first version of the manuscript. SL generated and contributed 50 K SNP data for sheep breeds in the Xinjiang biota. JZ, SL, MT, and C-lZ wrote and/or revised the manuscript. All authors read and approved the manuscript.

Funding

This study was funded by grants form National Natural Science Foundation of China (32060743) supported by Bintuan Science and Technology Program (2022CB001-09).

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.2023.1092066/full#supplementary-material

References

Akey, J. M., Zhang, G., Zhang, K., Jin, L., and Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814. doi:10.1101/gr.631202

Al-Mamun, H. A., Clark, S. A., Kwan, P., and Gondro, C. (2015). Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genet. Sel. Evol.* 47, 90. doi:10.1186/s12711-015-0169-6

Al-Samerria, S., Al-Ali, I., McFarlane, J. R., and Almahbobi, G. (2015). The impact of passive immunisation against BMPRIB and BMP4 on follicle development and ovulation in mice. *Reproduction* 149, 403–411. doi:10.1530/rep-14-0451

Aoto, K., Kato, M., Akita, T., Nakashima, M., Mutoh, H., Akasaka, N., et al. (2021). ATP6v0a1 encoding the a1-subunit of the v0 domain of vacuolar h^+-ATPases is essential for brain development in humans and mice. *Nat. Commun.* 12, 2107. doi:10.1038/s41467-021-22389-5

Bizjak, N., Tansek, M. Z., Stefanija, M. A., Lampret, B. R., Mezek, A., Torkar, A. D., et al. (2020). Precocious puberty in a girl with 3-methylglutaconic aciduria type 1 (3-MGA-i) due to a novel AUH gene mutation. *Mol. Genet. Metabolism Rep.* 25, 100691. doi:10.1016/j.ymgmr.2020.100691

Brackett, C. M., Greene, K. F., Aldrich, A. R., Trageser, N. H., Pal, S., Molodtsov, I., et al. (2021). Signaling through TLR5 mitigates lethal radiation damage by neutrophil-dependent release of MMP-9. *Cell. Death Discov.* 7, 266. doi:10.1038/s41420-021-00642-6

Buratini, J., Pinto, M., Castilho, A., Amorim, R., Giometti, I., Portela, V., et al. (2007). Expression and function of fibroblast growth factor 10 and its receptor, fibroblast growth factor receptor 2B, in bovine follicles. *Biol. Reproduction* 77, 743–750. doi:10.1095/biolreprod.107.062273

Caixeta, E. S., Sutton-McDowall, M. L., Gilchrist, R. B., Thompson, J. G., Price, C. A., Machado, M. F., et al. (2013). Bone morphogenetic protein 15 and fibroblast growth factor 10 enhance cumulus expansion, glucose uptake, and expression of genes in the ovulatory cascade during *in vitro* maturation of bovine cumulus–oocyte complexes. *Reproduction* 146, 27–35. doi:10.1530/rep-13-0079

Calder, P. C. (2015). Functional roles of fatty acids and their effects on human health. *J. Parenter. Enter. Nutr.* 39, 18S–32S. doi:10.1177/0148607115595980

Chen, J., Wang, Z., Guo, X., Li, F., Wei, Q., Chen, X., et al. (2019). TRIM66 reads unmodified h3r2k4 and h3k56ac to respond to DNA damage in embryonic stem cells. *Nat. Commun.* 10, 4273. doi:10.1038/s41467-019-12126-4

Chen, Y., Li, Y., Peng, Y., Zheng, X., Fan, S., Yi, Y., et al. (2018). *δnp63α* down-regulates c-myc modulator mm1 via e3 ligase herc3 in the regulation of cell senescence. *Cell. Death Differ.* 25, 2118–2129. doi:10.1038/s41418-018-0132-5

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., and Woolliams, J. A. (2010). Linkage disequilibrium and historical effective population size in the thoroughbred horse. *Anim. Genet.* 41, 8–15. doi:10.1111/j.1365-2052.2010.02092.x

Das, U. (2006). "Essential fatty acids," in *Encyclopedia of biophysics* (Heidelberg: Springer Berlin Heidelberg), 706–714. doi:10.1007/978-3-642-16712-6_533

Ding, F., Luo, X., Tu, Y., Duan, X., Liu, J., Jia, L., et al. (2021). Alpk1 sensitizes pancreatic beta cells to cytokine-induced apoptosis via upregulating TNF-*α* signaling pathway. *Front. Immunol.* 12, 705751. doi:10.3389/fimmu.2021.705751

Edea, Z., Dadi, H., Kim, S.-W., Park, J.-H., Shin, G.-H., Dessie, T., et al. (2014). Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. *J. Animal Breed. Genet.* 131, 358–366. doi:10.1111/jbg.12083

Fischer, C. D., Wachoski-Dark, G. L., Grant, D. M., Bramer, S. A., and Klein, C. (2018). Interferon epsilon is constitutively expressed in equine endometrium and up-regulated during the luteal phase. *Animal Reproduction Sci.* 195, 38–43. doi:10.1016/j.anireprosci.2018.05.003

Ghoreishifar, S. M., Moradi-Shahrbabak, H., Parna, N., Davoudi, P., and Khansefid, M. (2019). Linkage disequilibrium and within-breed genetic diversity in iranian zandi sheep. *Arch. Anim. Breed.* 62, 143–151. doi:10.5194/aab-62-143-2019

Grossi, D. A., Jafarikia, M., Brito, L. F., Buzanskas, M. E., Sargolzaei, M., and Schenkel, F. S. (2017). Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in canadian pigs. *BMC Genet.* 18, 6. doi:10.1186/s12863-017-0473-y

Hanson, L. Å., and Korotkova, M. (2002). The role of breastfeeding in prevention of neonatal infection. *Seminars Neonatol.* 7, 275–281. doi:10.1016/s1084-2756(02)90124-7

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

Hill, W. G. (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33, 229–239. doi:10.1038/hdy.1974.89

Irving-Rodgers, H., and Rodgers, R. (2006). Extracellular matrix of the developing ovarian follicle. *Seminars Reproductive Med.* 24, 195–203. doi:10.1055/s-2006-948549

Jiang, F., Zhu, Y., Chen, Y., Tang, X., Liu, L., Chen, G., et al. (2021). Progesterone activates the cyclic AMP-protein kinase a signalling pathway by upregulating ABHD2 in fertile men. *J. Int. Med. Res.* 49, 300060521999527. doi:10.1177/0300060521999527

Kijas, J. W., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., et al. (2014). Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Anim. Genet.* 45, 754–757. doi:10.1111/age.12197

Kumar, S., Rajput, P. K., Bahire, S. V., Jyotsana, B., Kumar, V., and Kumar, D. (2020). Differential expression of BMP/SMAD signaling and ovarian-associated genes in the granulosa cells of FecB introgressed GMM sheep. *Syst. Biol. Reproductive Med.* 66, 185–201. doi:10.1080/19396368.2019.1695977

Lamuedra, A., Gratal, P., Calatrava, L., Ruiz-Perez, V. L., Palencia-Campos, A., Portal-Núñez, S., et al. (2022). Blocking chondrocyte hypertrophy in conditional evc knockout mice does not modify cartilage damage in osteoarthritis. *FASEB J.* 36, e22258. doi:10.1096/fj.202101791rr

Lawrence, R. M., and Lawrence, R. A. (2004). Breast milk and infection. *Clin. Perinatology* 31, 501–528. doi:10.1016/j.clp.2004.03.019

Liu, S., He, S., Chen, L., Li, W., Di, J., and Liu, M. (2017). Estimates of linkage disequilibrium and effective population sizes in Chinese merino (xinjiang type) sheep by genome-wide SNPs. *Genes. and Genomics* 39, 733–745. doi:10.1007/s13258-017-0539-2

Lv, X., Chen, L., He, S., Liu, C., Han, B., Liu, Z., et al. (2020). Effect of nutritional restriction on the hair follicles development and skin transcriptome of Chinese merino sheep. *Animals* 10, 1058. doi:10.3390/ani10061058

Majkowski, M., Laszkiewicz, A., Sniezewski, L., Grzmil, P., Pawlicka, B., Tomczyk, I., et al. (2018). Lack of NWC protein (c11orf74 homolog) in murine spermatogenesis results in reduced sperm competitiveness and impaired ability to fertilize egg cells *in vitro*. *PLOS ONE* 13, e0208649. doi:10.1371/journal.pone.0208649

Mastrangelo, S., Portolano, B., Gerlando, R. D., Ciampolini, R., Tolone, M., Sardina, M., et al. (2017). Genome-wide analysis in endangered populations: A case study in barbaresca sheep. *Animal* 11, 1107–1116. doi:10.1017/s1751731116002780

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., et al. (2001). Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in booroola mérino ewes. *Proc. Natl. Acad. Sci.* 98, 5104–5109. doi:10.1073/pnas.091577598

Ospina, A. M. T., Maiorano, A. M., Curi, R. A., Pereira, G. L., Zerlotti-Mercadante, M. E., Cyrillo, J. N. S. G., et al. (2019). Linkage disequilibrium and effective population size in gir cattle selected for yearling weight. *Reproduction Domest. Animals* 54, 1524–1531. doi:10.1111/rda.13559

Pan, Y., Wang, M., Baloch, A. R., Zhang, Q., Wang, J., Ma, R., et al. (2019). FGF10 enhances yak oocyte fertilization competence and subsequent blastocyst quality and regulates the levels of CD9, CD81, DNMT1, and DNMT3b. *J. Cell. Physiology* 234, 17677–17689. doi:10.1002/jcp.28394

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi:10.1534/genetics.112.145037

Pinto, R. P., Fontes, P., Loureiro, B., Castilho, A. S., Ticianelli, J. S., Razza, E. M., et al. (2014). Effects of FGF10 on bovine oocyte meiosis progression, apoptosis, embryo development and relative abundance of developmentally important genes *in vitro*. *Reproduction Domest. Animals* 50, 84–90. doi:10.1111/rda.12452

Porto-Neto, L. R., Kijas, J. W., and Reverter, A. (2014). The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* 46, 22. doi:10.1186/1297-9686-46-22

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

Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2009). The pattern of linkage disequilibrium in German holstein cattle. *Anim. Genet.* 41, 346–356. doi:10.1111/j.1365-2052.2009.02011.x

Rao, Y. S., Liang, Y., Xia, M. N., Shen, X., Du, Y. J., Luo, C. G., et al. (2008). Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* 145, 251–257. doi:10.1111/j.1601-5223.2008.02043.x

Rodgers, R. J., and Irving-Rodgers, H. F. (2010). Formation of the ovarian follicular antrum and follicular fluid. *Biol. Reproduction* 82, 1021–1029. doi:10.1095/biolreprod.109.082941

Scotland, K. B., Chen, S., Sylvester, R., and Gudas, L. J. (2009). Analysis of rex1 (zfp42) function in embryonic stem cell differentiation. *Dev. Dyn.* 238, 1863–1877. doi:10.1002/dvdy.22037

Seira, O., Liu, J., Assinck, P., Ramer, M., and Tetzlaff, W. (2019). KIF2a characterization after spinal cord injury. *Cell. Mol. Life Sci.* 76, 4355–4368. doi:10.1007/s00018-019-03116-2

Singh, A., Kumar, A., Mehrotra, A., Karthikeyan, A., Pandey, A. K., Mishra, B. P., et al. (2021). Estimation of linkage disequilibrium levels and allele frequency distribution in crossbred vrindavani cattle using 50k SNP data. *PLOS ONE* 16, e0259572. doi:10.1371/journal.pone.0259572

Su, Z., Li, Y., Lv, H., Cui, X., Liu, M., Wang, Z., et al. (2021). CLEC14a protects against podocyte injury in mice with adriamycin nephropathy. *FASEB J.* 35, e21711. doi:10.1096/fj.202100283r

Sun, H.-M., Chen, X.-L., Chen, X.-J., Liu, J., Ma, L., Wu, H.-Y., et al. (2017). PALLD regulates phagocytosis by enabling timely actin polymerization and depolymerization. *J. Immunol.* 199, 1817–1826. doi:10.4049/jimmunol.1602018

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17, 520–526. doi:10.1101/gr.6023607

Terhorst, J., Kamm, J. A., and Song, Y. S. (2016). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* 49, 303–309. doi:10.1038/ng.3748

Uimari, P., and Tapio, M. (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J. Animal Sci.* 89, 609–614. doi:10.2527/jas.2010-3249

Valisno, J. A. C., May, J., Singh, K., Helm, E. Y., Venegas, L., Budbazar, E., et al. (2021). BCL11b regulates arterial stiffness and related target organ damage. *Circulation Res.* 128, 755–768. doi:10.1161/circresaha.120.316666

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72. doi:10.1371/journal.pbio.0040072

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Trans. R. Soc. B Biol. Sci.* 360, 1395–1409. doi:10.1098/rstb.2005.1682

Yang, Q. E., Giassetti, M. I., and Ealy, A. D. (2011). Fibroblast growth factors activate mitogen-activated protein kinase pathways to promote migration in ovine trophoblast cells. *Reproduction* 141, 707–714. doi:10.1530/rep-10-0541

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics Bioinforma.* 19, 619–628. doi:10.1016/j.gpb.2020.10.007

Yokoyama, S., Takahashi, A., Kikuchi, R., Nishibu, S., Lo, J. A., Hejna, M., et al. (2021). SOX10 regulates melanoma immunogenicity through an IRF4–IRF1 axis. *Cancer Res.* 81, 6131–6141. doi:10.1158/0008-5472.can-21-2078

Yuan, J., Luo, K., Deng, M., Li, Y., Yin, P., Gao, B., et al. (2014). HERC2-USP20 axis regulates DNA damage checkpoint through claspin. *Nucleic Acids Res.* 42, 13110–13121. doi:10.1093/nar/gku1034

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2018). 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

Zhang, K., Hansen, P. J., and Ealy, A. D. (2010). Fibroblast growth factor 10 enhances bovine oocyte maturation and developmental competence *in vitro*. *Reproduction* 140, 815–826. doi:10.1530/rep-10-0190

Zhang, X., Zhang, L., Sun, W., Lang, X., Wu, J., Zhu, C., et al. (2020). Study on the correlation between BMPR1b protein in sheep blood and reproductive performance. *J. Animal Sci.* 98, skaa100. doi:10.1093/jas/skaa100

Zhu, B., Chen, C., Xue, G., Lei, X., Li, J., Moyzis, R. K., et al. (2014). The GABRB1 gene is associated with thalamus volume and modulates the association between thalamus volume and intelligence. *NeuroImage* 102, 756–763. doi:10.1016/j.neuroimage.2014.08.048

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fvets.2023.1134903

Check for updates

OPEN ACCESS

EDITED BY
Ibrar Muhammad Khan,
Fuyang Normal University, China

REVIEWED BY
Cemal Ün,
Ege University, Türkiye
Muhammad Zahoor,
University of Agriculture, Dera Ismail Khan, Pakistan

*CORRESPONDENCE
Hongwei Xu
✉ xuhongwei@xbmu.edu.cn
Xianyong Lan
✉ lanxianyong79@nwsuaf.edu.cn

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

RECEIVED 31 December 2022
ACCEPTED 27 March 2023
PUBLISHED 17 April 2023

CITATION
Luo Y, Akhatayeva Z, Mao C, Jiang F, Guo Z, Xu H and Lan X (2023) The ovine *HIAT1* gene: mRNA expression, InDel mutations, and growth trait associations. *Front. Vet. Sci.* 10:1134903. doi: 10.3389/fvets.2023.1134903

COPYRIGHT
© 2023 Luo, Akhatayeva, Mao, Jiang, Guo, Xu and Lan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The ovine *HIAT1* gene: mRNA expression, InDel mutations, and growth trait associations

Yunyun Luo[1], Zhanerke Akhatayeva[1], Cui Mao[1,2], Fugui Jiang[2], Zhengang Guo[3], Hongwei Xu[4]* and Xianyong Lan[1]*

[1]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China, [2]Shandong Key Lab of Animal Disease Control and Breeding, Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan, China, [3]Bijie Animal Husbandry and Veterinary Science Research Institute, Bijie, China, [4]College of Life Science and Engineering, Northwest Minzu University, Lanzhou, China

Background: The *hippocampal abundant transcript 1 (HIAT1)* gene, also known as *major facilitator superfamily domain-containing 14A (MFSD14A)*, encodes for a transmembrane transporter protein and has been previously shown to be associated with milk production in buffalo and sheep breeds, as well as growth traits in chicken and goats. However, tissue level distribution of the ovine HIAT1 gene, as well as its effect on body morphometric traits in sheep, has yet to be studied.

Methods: The *HIAT1* mRNA expression profile of Lanzhou fat-tailed (LFT) sheep was determined by quantitative real-time PCR (qPCR). A total of 1498 sheep of three indigenous Chinese sheep breeds were PCR-genotyped for polymorphisms of *HIAT1* gene. Student's t-test was used to observe the association between the genotype and sheep morphometric traits.

Results: *HIAT1* was widely expressed in all examined tissues, and was particularly abundant in the testis of male LFT sheep. Additionally, a 9-bp insertion mutation (rs1089950828) located within the 5'-upstream region of *HIAT1* was investigated in Luxi black-headed (LXBH) sheep and Guiqian semi-fine wool (GSFW) sheep. The wildtype allele frequency 'D' was found to be more prevalent than that of the mutant allele 'I'. Furthermore, low genetic diversity was confirmed in all sampled sheep populations. Subsequent association analyses indicated an association between the 9-bp InDel mutation of interest and the morphometric traits of LXBH and GSFW sheep. Furthermore, yearling ewes with a heterozygous genotype (ID) demonstrated smaller body sizes, while yearling rams and adult ewes with the heterozygous genotype were found to have overall better growth performance.

Conclusion: These findings imply that functional InDel polymorphism (rs1089950828) has the potential to be utilized for marker-assisted selection (MAS) of growth traits in domestic Chinese sheep populations.

KEYWORDS

sheep, HIAT1 gene, insertion/deletion (InDel), growth traits, marker-assisted selection (MAS)

1. Introduction

The final grown performance of sheep and other domesticated animals is affected by multiple factors including environmental, nutritional, and genetic influences (1). Using high-throughput whole genome sequencing, RNA sequencing, and molecular biology techniques, a number of candidate functional genes and genetic variants linked to important growth traits in livestock were found and used for the purposes of molecular-assisted and omics-assisted breeding (2–6). Among the numerous genome variants including SNPs

(single nucleotide polymorphisms), CNVs (copy number variations), and other genetic markers, numerous studies have demonstrated that InDels (insertion/deletion) function by regulating gene translation or altering a protein and are related to differences in the characteristics of individual animals (7–9).

The *hippocampal abundant transcript 1* (*HIAT1*) gene is also known as the *major facilitator superfamily domain 14A* (*MFSD14A*) gene (10). The HIAT1 protein contains a conserved sugar transporter motif as well as 12 transmembrane alpha helices (TMs) linked by hydrophilic loops, and belongs to the major facilitator superfamily (MFS) of secondary protein transporters (11). The MFS protein superfamily is the largest secondary transporter family present in a broad variety of species from archaea to mammals and plays an essential role in the exchange of cellular materials and energy metabolism, which facilitates the transport of many substrates across both cytoplasmic and internal cell membranes (12). The expression of both *MFSD1* and *MFSD3* genes has previously been shown to be significantly different in both the cerebellum and brain of mice after they were subjected to starvation and high-fat diets. This suggests that these genes are regulated at the nutrient level and may be related to nutrient intake and transport (13). Two members of the MFS gene family, *MFSD6* and *MFSD10* were found to be related to the energy metabolism of cells. Expression levels of these two genes were found to be down-regulated when the cellular energy consumption rate increased or when cellular energy was significantly depleted (14). These findings suggest that *HIAT1* may also be related to fatty acid uptake and transport.

Chinese scholars have found that *HIAT1* affects buffalo milk fat synthesis mainly through regulation of cellular membrane transport functions (15). Moreover, studies have shown that the HIAT1 protein has the capacity to transport solute from the blood stream that is required for spermiogenesis, thus disruption of this gene has been shown to cause round-headed sperm and sterility in male mice (16). *HIAT1* has previously been shown to be commonly expressed in multiple tissues types and highly expressed in testis, suggesting that it has extensive biological functions (17). *HIAT1* is expressed in brown adipose tissue and is negatively correlated with the process in which glucose is converted to fat (18). Additionally, polymorphisms in *HIAT1* have been shown to affect milk traits in buffalo as well as growth in goats (15, 19). However, to date, there are few studies that have reported on the function of *HIAT1* in indigenous Chinese sheep populations.

In order to further improve the production features of native Chinese sheep populations, the current study aimed to detect *HIAT1* expression features in Lanzhou fat-tailed (LFT) sheep, explore InDel variations of the *HIAT1* gene in three indigenous Chinese sheep breeds, which included Luxi black-headed (LXBH) sheep from east China, Guiqian semi-fine wool (GSFW) sheep from southwest China, and LFT sheep from northwest China, as well as conduct association analyses between sheep genotypes and growth traits to provide a basis for use of the *HIAT1* gene for sheep selection and breeding.

2. Materials and methods

2.1. Animals and ethics statement

Twelve tissue samples were collected from three 6-month-old male Lanzhou fat-tailed sheep and include heart, liver, spleen, lung, kidney, testis, longissimus dorsi muscle, tail adipose, perirenal fat, subcutaneous fat, small intestine, and rennet stomach (Yongjing county, Gansu Province). A total of 1498 indigenous Chinese sheep were used for DNA experiments and included Luxi black-headed sheep ($n = 616$) raised in Liaocheng city, Shandong province; Guiqian semi-fine wool sheep ($n = 824$) raised in Bijie city, Guizhou province; and Lanzhou fat-tailed sheep ($n = 58$) provided by Gansu Ruilin Science and Technology Breeding Company (Yongjing county, Gansu province). All tissue samples were frozen in liquid nitrogen and stored at −80°C for RNA and DNA isolation. Relevant body morphometric traits were recorded, including body weight, chest circumference, height at hip cross, body height, body length, cannon circumference, hip width, abdominal circumference, chest depth, and chest width. Animal experiments were conducted in line with the guidelines for the care and use of animals and approved by the Animal Ethics Committee of Northwest A&F University.

2.2. DNA isolation, primer design, and InDel genotyping

Genomic DNA was isolated following previously established protocols (20, 21). Total RNA and genomic DNA concentrations were determined using Nanodrop 2000. DNA was diluted to a

TABLE 1 Primers used for InDel detection and qPCR analysis of ovine *HIAT1*.

Names	ID	Alleles	Primer sequences (5′ to 3′)	Sizes (bp)
P1-ins-5bp	rs598326621	-/TTAAG	F: TTCCTTCACTCCTTAAGACTTCG R: TTTGATTGTGATGACTGTACTGT	89/84
P2-del-5bp	rs604922868	TCAGT/-	F: CGCAATTCCCCCATTAAAT R: GCAAAGAGTCGGACACGAC	109/104
P3-ins-9bp	rs1089950828	-/GTCCAGTGG	F: TTCCCTGTTCATCACCAACTC R: ACCTTTTCTTTATTCCCTGCC	148/139
HIAT1	XM_004002229.5	-	F: TACTGCTGGCTCTGCTTGTTGC R: TACTGTGAGGATGGCTGTGACTACC	108
GAPDH	NM_001190390.1	-	F: CCTGCCAAGTATGATGAGAT R: TGAGTGTCGCTGTTGAAGT	119

FIGURE 1
Amino acid sequence alignment of *HIAT1* in different mammalian species. The red boxes denote the parts of the sequence that are unable to be matched with the *Ovis aries* sequence. Note: • denotes the same amino acid.

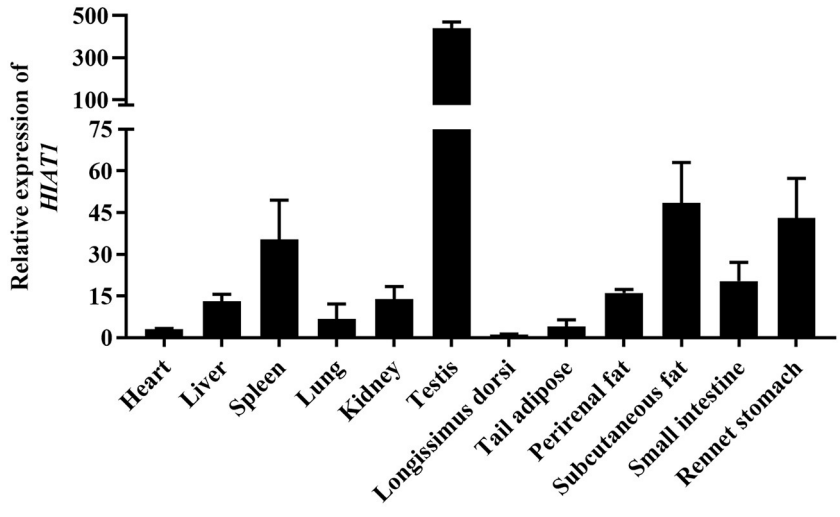


FIGURE 2
HIAT1 mRNA expression patterns detected by qPCR in Lanzhou fat-tailed sheep.

concentration of 20 ng/μL and stored at −40°C. Thirty DNA samples were randomly selected from each breed to construct three genomic DNA pools for investigation of *HIAT1* variations.

Three upstream gene variants were screened from the Ensembl database, and then several primers (Table 1) were designed using Primer Premier 5 software and synthesized by Sangon Biotech (Xi'an, China) based on reference sequences in ovine *HIAT1* (NC 056054.1). Polymorphic fragments were amplified using the polymerase chain reaction (PCR) method and utilized a touch-down program and reaction volumes as described in

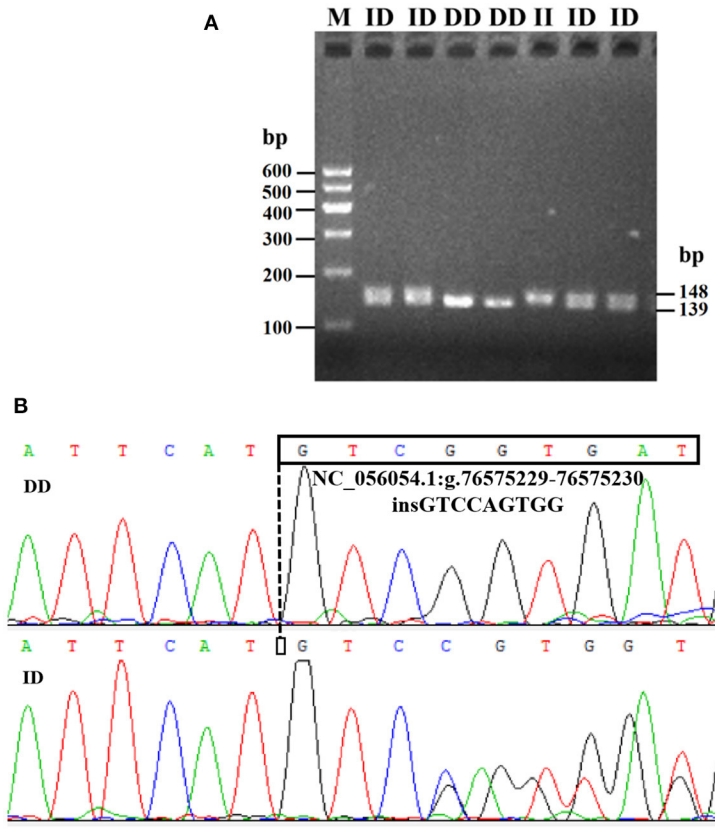


FIGURE 3
Electrophoresis gel **(A)** and sequence chromatograms **(B)** of the P3-ins-9bp InDel within the ovine *HIAT1* gene.

previous studies (8, 22). PCR products were then separated by electrophoresis on a 3.5% agarose gel stained with GoldView. The genotype of each individual sheep was analyzed.

2.3. RNA isolation, cDNA synthesis, and quantitative real-time PCR

Total RNA was extracted from tissue samples using TRIzol reagent, and cDNA was synthesized according to manufactures' instructions using an Evo M-MLV RT Kit with gDNA Clean for qPCR II (Accurate biology Co., Ltd, Changsha, China). A LightCycler 96 real-time PCR system (Roche, Switzerland) was used to run the qPCR program. qPCR reactions (10 μL) contained 5 μL 2 × ChamQ SYBR qPCR Master Mix (Vazyme, Nanjing, China), 0.2 μL of each primer, and 4.6 μL cDNA (1/100 dilution). qPCR conditions were as follows: 95°C for 180 s, followed by 40 amplification cycles lasting 40 s each (95°C for 10 s and 60°C for 30 s). Gene expression was quantitated using the $2^{-\Delta\Delta Ct}$ method.

2.4. Bioinformatics and statistical analysis

Differing amino acid sequences alignment were made using DNAMAN 6.0 software. The percent similarity of amino acid sequences was analyzed by NCBI-Blast online software. Using AliBaba2.1 (http://gene-regulation.com/pub/programs/alibaba2/), predictions of transcription factor binding to the mutant ovine *HIAT1* gene sequence were generated (23). Population genetic indices, as well as allelic and genotypic frequencies, were determined. χ^2 test was performed to determine whether the gene variants are within Hardy-Weinberg equilibrium (HWE) and analyze the difference in genotype distributions amongst each sheep population. The statistical model used was $Y_{ij} = \mu + G_i + \varepsilon_{ij}$, where Y_{ij} = observed growth traits, μ = population mean, G_i = fixed effect of the genotype, ε_{ij} = random error. If there were only two genotypes, student's *t*-test was performed to determine the effect of different genotypes on economic traits.

3. Results

3.1. Multiple sequence alignment and mRNA expression

The results of NCBI-BLASTP sequence analysis showed that the *HIAT1* in *Ovis aries* (Genbank: XP_004002278.1) shares 100%, 100%, 99.80%, 99.80% and 99.59% similarity with that of *Capra hircus* (XP_017901305.1), *Bos taurus* (NP_001095508.1), *Mus musculus* (NP_032272.2), *Sus scrofa* (XP_013852805.2) and *Homo sapiens* (NP_149044.2), respectively (Figure 1). Additionally,

10.3389/fvets.2023.1134903

TABLE 2 Genotypic frequencies, allelic frequencies, and genetic diversity of the P3-ins-9bp InDel.

Breeds (n)	Genotypic frequencies			Allelic frequencies		HWE *P*-value	Population parameters		
	II (n)	ID (n)	DD (n)	I	D		He	Ne	PIC
LXBH (616)	0.006 (4)	0.156 (96)	0.838 (516)	0.084	0.916	*P*>0.05	0.155	1.183	0.143
GSFW (824)	0.004 (3)	0.080 (66)	0.916 (755)	0.044	0.956	*P*>0.05	0.084	1.091	0.080
LFT (58)	0	0	1.000 (58)	0	1.000	-	0	1.000	0

LXBH, Luxi black-headed sheep; GSFW, Guiqian semi-fine wool sheep; LFT, Lanzhou fat-tailed sheep; II, insertion/insertion; ID, insertion/deletion; DD, deletion/deletion; HWE, Hardy-Weinberg equilibrium; He, heterozygosity; Ne, effective allele numbers; PIC, Polymorphism information content.

TABLE 3 Chi-square test of *HIAT1* P3-ins-9bp genotype distribution in the three sheep breeds.

Breeds	LXBH	GSFW	LFT
LXBH		7e-6**	8.6e-5**
GSFW	1.4e-5**		0.013*
LFT	4.24e-4**	0.038*	

Results below the diagonal line represent the Chi-square test of genotype frequency for each sheep breed; Results above the diagonal line represent the Chi-square test of allele frequency for each sheep breed. $^{*}P < 0.05$, $^{**}P < 0.01$.

qPCR results revealed that *HIAT1* was expressed in most sheep tissues, especially in the testis of male LFT sheep (Figure 2), which coincided with expression and distribution in human tissues.

3.2. Detection and genetic parameter analysis of InDel polymorphisms within the *HIAT1* gene

Electrophoresis patterning and sequence mapping of the P3-ins-9bp InDel showed polymorphisms in the 5'-upstream region of *HIAT1* for both Luxi black-headed sheep and Guiqian semi-fine wool sheep. There were no detected *HIAT1* polymorphisms found in Lanzhou fat-tailed sheep (Figure 3). Genotyping analyses showed that the *HIAT1* P3-ins-9bp locus contained three genotypes (the homozygous insertion type, II, 148 bp; the homozygous deletion type, DD, 139 bp; and the heterozygous mutation type, ID, 148 bp and 139 bp) for both Luxi black-headed sheep and Guiqian semi-fine wool sheep; however, only one genotype (DD) was discovered in the Lanzhou fat-tailed breed.

The genotypic frequency, allelic frequency, and population genetic indices were subsequently calculated and shown in Table 2. The DD genotype frequency was found to be much higher than both ID and II genotypes. Amongst all sheep breeds tested, the P3-ins-9bp locus was found to have low genetic polymorphism ($PIC<0.25$) while maintaining Hardy-Weinberg equilibrium ($P > 0.05$). χ^2 testing revealed that there were significant differences both in the genotypic and allelic frequency distributions of the P3-ins-9bp locus amongst all three populations ($P < 0.05$) (Table 3).

3.3. Association between the P3-ins-9bp InDel and growth traits

The associations between the P3-ins-9bp InDel of the ovine *HIAT1* gene and growth performance were analyzed (Table 4). For LXBH sheep, t-tests results confirmed that the P3-ins-9bp InDel had no remarkable influence on growth traits for lambs ($P > 0.05$) (not shown). Yearling rams with the homozygous DD genotype were found to have lower chest depth, body weight, cannon circumference, and body height ($P < 0.05$). Yearling ewes with the homozygous DD

TABLE 4 Association between the *HIAT1* P3-ins-9bp locus and sheep growth traits.

Breeds	Types (n)	Traits	Observed genotypes (LSM ± SE)		*P*-values
LXBH	Yearling rams		ID (9)	DD (70)	
	(79)	Body weight (kg)	68.78 ± 7.00	54.47 ± 2.02	0.023
		Body height (cm)	70.67 ± 1.76	66.53 ± 0.68	0.041
		Chest depth (cm)	31.11 ± 1.15	27.63 ± 0.54	0.029
		Cannon circumference (cm)	10.22 ± 0.21	9.54 ± 0.11	0.036
	Yearling ewes		ID (16)	DD (82)	
	(98)	Body weight (kg)	39.99 ± 3.10	48.92 ± 1.52	0.018
		Height at the hip cross (cm)	64.38 ± 1.19	66.94 ± 0.51	0.044
		Chest depth (cm)	24.60 ± 0.83	27.47 ± 0.45	0.010
		Chest circumference (cm)	80.06 ± 2.20	86.87 ± 1.05	0.009
		Abdominal circumference (cm)	100.81 ± 3.07	108.52 ± 1.21	0.013
	Adult ewes		ID (15)	DD (58)	
	(73)	Body height (cm)	71.13 ± 1.31	68.05 ± 0.47	0.009
		Height at the hip cross (cm)	72.13 ± 1.18	69.16 ± 0.45	0.006
		Cannon circumference (cm)	10.20 ± 0.17	9.33 ± 0.12	0.001
GSFW	Adult ewes		ID (19)	DD (155)	
	(174)	Body weight (kg)	55.26 ± 2.00	49.81 ± 0.50	0.001
		Body length (cm)	82.68 ± 1.07	80.69 ± 0.31	0.037
		Chest width (cm)	30.05 ± 3.01	28.41 ± 0.18	0.005
		Chest circumference (cm)	99.63 ± 1.75	95.17 ± 0.49	0.004

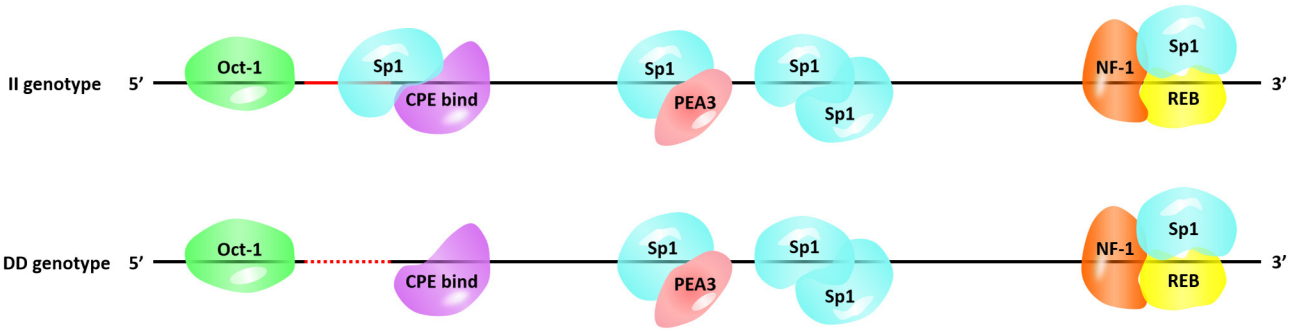


FIGURE 4
Transcription factor binding sites on the 9-bp InDel locus.

genotype had significantly higher body weight, chest depth, chest circumference, height at the hip cross and abdominal circumference than those with the heterozygous ID genotype ($P < 0.05$). Additionally, adult ewes exhibited significant differences in height at the hip cross, body height, and cannon circumference ($P < 0.01$), with the heterozygous ID genotype being the dominant individual. For GSFW sheep, adult ewes with the heterozygous ID genotype showed superior performance compared to those with the homozygous DD genotype with regards to body weight, body length, chest width, and chest circumference.

4. Discussion

Recently, a study by Liu et al. (24) identified a quantitative trait locus (QTL) that affects milk traits in buffalo breed and includes *SLC35A3* (*Solute carrier family 35 member A3*) and *HIAT1* (24). The *SLC35A3* gene is located adjacent to the *HIAT1* gene and is involved in the transport of glucose, vitamins and other substances (25). A genome-wide association study (GWAS) on milk production traits determined that *SLC35A3* may also play a significant role in sheep milk traits (26). Moreover, a missense mutation in bovine *SLC35A3* has been found to lead to complex

vertebral malformation (25). A recent study by Ye (15) found that *HIAT1* regulates milk fat synthesis in mammary epithelial cells through the PPAR signaling pathway, and that two SNPs within *HIAT1* are significantly associated with both milk fat and milk protein rates in buffalo (15). A more surprising finding is that both *HIAT1* and *SLC35A3* have the potential to affect both carcass and growth traits in chickens (27). In this study, we found that *HIAT1* has high amino acid sequence conservation, and that its tissue expression pattern is similar to that of humans. These results suggest that *HIAT1* might have a similar function amongst multiple types of mammals. Moreover, *HIAT1* mRNA expression was detected in perirenal fat, subcutaneous fat, and the livers of LFT sheep, suggesting that *HIAT1* may be related to fat deposition in sheep. These finding suggest *HIAT1* is a key candidate gene for regulating important economic traits in livestock.

Body morphology traits reflect the growth and development of different parts of the body and are also important economic and breeding indicators for both meat and meat-wool sheep breeds (28, 29). Existing studies have showed that the *HIAT1* gene influenced both growth and meat production traits in goats, and that a 15-bp insertion in *HIAT1* was associated with body morphology in Shaanbei white cashmere goats (19, 30). In this study, a 9-bp insertion mutation located in the 5'-upstream region of the ovine *HIAT1* gene (rs1089950828 -/GTCCAGTGG) was elucidated. It is well known that variations in the 5'-upstream region of genes has the potential to affect gene transcription and mRNA expression levels through changing transcription factor binding sites (TFBSs) and can also influence an individual's phenotype (31, 32). Thus, this research studied the association between the P3-ins-9bp InDel and sheep growth traits. Furthermore, the P3-ins-9bp InDel was strongly associated with body measurement traits in both LXBH and GSFW sheep based on association analysis. For yearling LXBH sheep, ewes with a heterozygous ID genotype demonstrated overall lower body size, while rams with a heterozygous ID genotype had better growth performance. These results indicate that the 9-bp insertion mutation may have a negative regulatory effect on both growth and development of LXBH yearling ewes as well as promote growth in yearling rams. Additionally, a heterozygous ID genotype in ewes was found to inhibit body size development in yearling sheep as well as promote growth in adult sheep. This may be due to the effect of the 9-bp InDel on the differential expression of *HIAT1* at different growth stages throughout the sheep lifecycle (33). Furthermore, the PIC value is an indicator of molecular marker quality in genetic studies. Our InDel locus of interest showed low polymorphism ($PIC < 0.25$) amongst all tested sheep according to the PIC value, indicating that there was sufficient space in which artificial selection can take place (34, 35).

Next, we searched for the putative transcription factor binding sites that contained our InDel of interest. The P3-ins-9bp locus was found 624-bp upstream of the *HIAT1* transcriptional start site (TSS), which created an additional binding site for the transcription factor Sp1 (Specificity protein 1) falling upstream of the four existing Sp1 binding sites (Figure 4). SP1 is a transcription factor of the SP/KLF family. The regulation of SP1 with regards to target genes is achieved by binding of the DNA binding domain to GC-rich motifs in the promoter region of the target gene, thereby regulating cellular processes such as autophagy, apoptosis, proliferation, differentiation and angiogenesis (36, 37). Chen et al. (38) found that Sp1 binding sites in the core promoter region are essential for positive regulation of *FGF21* (*Fibroblast growth factor 21*) through gene transcription in both hepatic and adipose tissue (38). Furthermore, a recent study found that an SNP g.133A>C located within the SCD (*Stearoyl CoA desaturase*) promoter generated overall higher Sp1 binding affinity to the SCD promoter, which consequently affected gene expression (39). Moreover, a T > C mutation in the promoter region of *IGF1* (*Insulin-like growth factor 1*) was significantly associated with the litter size of Yunshang black goats. This mutation was found to create a new binding site for SP1, thereby promoting goat granulosa cell proliferation by regulating the expression of *IGF1* (40). Therefore, the interaction between Sp1 and *HIAT1* leads to changes in gene expression at both the mRNA and protein levels, which in turn has the potential to affect the growth traits of sheep. In future studies this interaction is worthy of further in-depth research and exploration. Furthermore, the molecular mechanism of action and associated allelic consequences should be further validated in future studies.

5. Conclusion

This study investigated both the mRNA expression profile and genetic variations of the ovine *HIAT1* gene. Our results demonstrated a high mRNA expression level of *HIAT1* in the testis of Lanzhou fat-tailed sheep. Moreover, a 9-bp insertion mutation (rs1089950828) in the 5'-upstream region of ovine *HIAT1* was detected. The frequency of the mutant allele "I" was found to be low. Furthermore, the 9-bp InDel locus was found to be closely associated with growth traits of LXBH and GSFW sheep, indicating that this InDel has the potential to be used as a genetic marker for assisted selection programs in domestic sheep.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

All animal experiments were conducted in accordance with the the China national standard of Guidelines on Welfare and Ethical Review for Laboratory Animals (GB/T 35892-2018). Our study was approved by Institutional Animal Care and Use Committee (IACUC) of Northwest A&F University.

Author contributions

YL, XL, and HX came up with idea and revised the manuscript. YL wrote the manuscript and performed the experiments. CM, FJ, ZG, and HX collected the sheep samples and isolated the genomic DNA. YL and ZA analyzed the data. All authors approved the final version of the manuscript for submission, contributed to the article, and approved the submitted version.

Funding

This work was supported by the National Natural and Science Foundation of China (32060741).

Acknowledgments

We thank Shaanxi Key Laboratory of Molecular Biology for Agriculture and the Life Science Research Core Services (LSRCS) of Northwest A&F University (Northern Campus) for their cooperation and support. We would like to thank the staff of Shandong Academy of Agricultural Sciences, Guizhou Xinwumeng Ecological Animal Husbandry Development Co., Ltd., and Gansu Ruilin Science and Technology Breeding Company for providing samples and phenotypic traits data for us.

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

1. Li ZH Li H, Zhang H, Wang SZ, Wang QG, Wang YX. Identification of a single nucleotide polymorphism of the insulin-like growth factor binding protein 2 gene and its association with growth and body composition traits in the chicken. *J Anim Sci.* (2006) 84:2902–6. doi: 10.2527/jas.2006-144

2. Bakhtiarizadeh MR, Alamouti AA. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. *Sci Rep.* (2020) 10:13525. doi: 10.1038/s41598-020-70527-8

3. Gu B, Sun R, Fang X, Zhang J, Zhao Z, Huang D, et al. Genome-wide association study of body conformation traits by whole genome sequencing in Dazu Black Goats. *Animals.* (2022) 12:548. doi: 10.3390/ani12050548

4. Salek Ardestani S, Jafarikia M, Sargolzaei M, Sullivan B, Miar Y. Genomic prediction of average daily gain, back-fat thickness, and loin muscle depth using different genomic tools in Canadian Swine populations. *Front Genet.* (2021) 12:665344. doi: 10.3389/fgene.2021.665344

5. Wang K, Liu X, Qi T, Hui Y, Yan H, Qu L, et al. Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (*Capra hircus*). *Genomics.* (2021) 113:142–50. doi: 10.1016/j.ygeno.2020.11.024

6. Ren F, Yu S, Chen R, Lv XY, Pan CY. Identification of a novel 12-bp insertion/deletion (indel) of iPS-related Oct4 gene and its association with reproductive traits in male piglets. *Anim Reprod Sci.* (2017) 178:55–60. doi: 10.1016/j.anireprosci.2017.01.009

7. Gamarra D, Aldai N, Arakawa A, de Pancorbo MM, Taniguchi M. Effect of a genetic polymorphism in SREBP1 on fatty acid composition and related gene expression in subcutaneous fat tissue of beef cattle breeds. *Anim Sci J.* (2021) 92:e13521. doi: 10.1111/asj.13521

8. Wang Z, Zhang X, Jiang E, Yan H, Zhu H, Chen H, et al. InDels within caprine IGF2BP1 intron 2 and the 3'-untranslated regions are associated with goat growth traits. *Anim Genet.* (2020) 51:117–21. doi: 10.1111/age.12871

9. 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

10. Fredriksson R, Nordström KJV, Stephansson O, Hägglund MGA, Schiöth HB. The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. *FEBS Lett.* (2008) 582:3811–6. doi: 10.1016/j.febslet.2008.10.016

11. Collins S, Martin TL, Surwit RS, Robidoux J. Genetic vulnerability to diet-induced obesity in the C57BL/6J mouse: physiological and molecular characteristics. *Physiol. Behav.* (2004) 81:243–8. doi: 10.1016/j.physbeh.2004.02.006

12. Dittami SM, Barbeyron T, Boyen C, Cambefort J, Collet G, Delage L, et al. Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi" Ec32, a new candidate genus of Alphaproteobacteria frequently associated with brown algae. *Front Genet.* (2014) 5:241. doi: 10.3389/fgene.2014.00241

13. Perland E, Hellsten SV, Lekholm E, Eriksson MM, Arapi V, Fredriksson R. The novel membrane-bound proteins MFSD1 and MFSD3 are putative SLC transporters affected by altered nutrient Intake. *J Mol Neurosci.* (2017) 61:199–214. doi: 10.1007/s12031-016-0867-8

14. Bagchi S, Perland E, Hosseini K, Lundgren J, Al-Walai N, Kheder S, et al. Probable role for major facilitator superfamily domain containing 6 (MFSD6) in the brain during variable energy consumption. *Int J Neurosci.* (2020) 130:476–89. doi: 10.1080/00207454.2019.1694020

15. Ye T. *Identification of Milk Fat Regulation Gene MFSD14A and Its Mechanism in Buffalo*. Wuhan: Huazhong Agricultural University. (2022).

16. Doran J, Walters C, Kyle V, Wooding P, Hammett-Burke R, Colledge WH. Mfsd14a (Hiat1) gene disruption causes globozoospermia and infertility in male mice. *Reproduction.* (2016) 152:91–9. doi: 10.1530/REP-15-0557

17. Sreedharan S, Stephansson O, Schioth HB, Fredriksson R. Long evolutionary conservation and considerable tissue specificity of several atypical solute carrier transporters. *Gene.* (2011) 478:11–8. doi: 10.1016/j.gene.2010.10.011

18. Pravenec M, Saba LM, Zídek V, Landa V, Mlejnek P, Šilhavý J, et al. Systems genetic analysis of brown adipose tissue function. *Physiol Genomics.* (2018) 50:52–66. doi: 10.1152/physiolgenomics.00091.2017

19. Gao JY, Song XY, Wu H, Tang Q, Wei ZY, Wang XY, et al. Detection of rs665862918 (15-bp Indel) of the HIAT1 gene and its strong genetic effects on growth traits in goats. *Animals.* (2020) 10:358–66. doi: 10.3390/ani10020358

20. Aljanabi SM, Martinez I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* (1997) 25:4692–3. doi: 10.1093/nar/25.22.4692

21. Lan XY, Pan CY, Chen H, Zhang CL Li JY, Zhao M, Lei CZ, et al. An AluI PCR-RFLP detecting a silent allele at the goat POU1F1 locus and its association with production traits. *Small Ruminant Res.* (2007) 73:8–12. doi: 10.1016/j.smallrumres.2006.10.009

22. Li X, Jiang E, Zhang K, Zhang S, Jiang F, Song E, et al. Genetic variations within the bovine CRY2 gene are significantly associated with carcass traits. *Animals.* (2022) 12:1616. doi: 10.3390/ani12131616

23. Zhang S, Kang Z, Sun X, Cao X, Pan C, Dang R, et al. Novel lncRNA lncFAM200B: molecular characteristics and effects of genetic variants on promoter activity and cattle body measurement traits. *Front Genet.* (2019) 10:968. doi: 10.3389/fgene.2019.00968

24. Liu JJ, Liang AX, Campanile G, Plastow G, Zhang C, Wang Z, et al. Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. *J Dairy Sci.* (2018) 101:433–44. doi: 10.3168/jds.2017-13246

25. Thomsen B, Horn P, Panitz F, Bendixen E, Petersen AH, Holm LE, et al. missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res.* (2006) 16:97–105. doi: 10.1101/gr.3690506

26. Rezvannejad E, Asadollahpour Nanaei H, Esmailizadeh A. Detection of candidate genes affecting milk production traits in sheep using whole-genome sequencing analysis. *Vet Med Sci.* (2022) 8:1197–204. doi: 10.1002/vms3.731

27. Zhang H, Shen LY, Xu ZC, Kramer LM Yu JQ, Zhang XY, Na W, et al. Haplotype-based genome-wide association studies for carcass and growth traits in chicken. *Poult Sci.* (2020) 99:2349–61. doi: 10.1016/j.psj.2020.01.009

28. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Selection Evol.* (2015) 47:66. doi: 10.1186/s12711-015-0142-4

29. Macé T, González-García E, Pradel J, Parisot S, Carrière F, Douls S, et al. Genetic analysis of robustness in meat sheep through body weight and body condition score changes over time. *J Anim Sci.* (2018) 96:4501–11. doi: 10.1093/jas/sky318

30. Zhang B, Chang L, Lan XY, Asif N, Guan FL, Fu DK, et al. Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (Capra species) breeds for the production of dairy, cashmere, and meat. *GigaScience*. (2018) 7:giy105. doi: 10.1093/gigascience/giy105

31. Glanzmann B, Lombard D, Carr J, Bardien S. Screening of two indel polymorphisms in the 5'UTR of the DJ-1 gene in South African Parkinson's disease patients. *J Neural Transm.* (2014) 121:135–8. doi: 10.1007/s00702-013-1094-x

32. Oner Y, Keskin A, Ustuner H, Soysal D, Karakaş V. Genetic diversity of the 3' and 5' untranslated regions of the HSP70. 1 gene between native Turkish and Holstein Friesian cattle breeds South African. *J Anim Sci.* (2017) 47:424–39. doi: 10.4314/sajas.v47i4.2

33. Cui Y, Yan H, Wang K, Xu H, Zhang X, Zhu H, et al. Insertion/Deletion within the KDM6A gene is significantly associated with litter size in goat. *Front Genet.* (2018) 9:91. doi: 10.3389/fgene.2018.00091

34. Serrote CML, Reiniger LRS, Silva KB, Rabaiolli SMDS, Stefanel CM. Determining the polymorphism information content of a molecular marker. *Gene.* (2020) 726:144175–8. doi: 10.1016/j.gene.2019.144175

35. Zhao HD, He S, Wang SH, Zhu YJ, Xu HW, Luo RY, et al. Two new insertion/deletion variants of the PITX2 gene and their effects on growth traits in sheep. *Anim Biotechnol.* (2018) 29:276–82. doi: 10.1080/10495398.2017.1379415

36. Guido C, Panza S, Santoro M, Avena P, Panno ML, Perrotta I, et al. Estrogen receptor beta (ERβ) produces autophagy and necroptosis in human seminoma cell line through the binding of the Sp1 on the phosphatase and tensin homolog deleted from chromosome 10 (PTEN) promoter gene. *Cell cycle.* (2012) 11:2911–21. doi: 10.4161/cc.21336

37. Vellingiri B, Iyer M, Devi Subramaniam M, Jayaramayya K, Siama Z, Giridharan B, et al. Understanding the role of the transcription factor Sp1 in ovarian cancer: from theory to practice. *Int J Mol Sci.* (2020) 21:1153. doi: 10.3390/ijms21031153

38. Chen S, Li H, Zhang J, Jiang S, Zhang M, Xu Y, et al. Identification of Sp1 as a transcription activator to regulate fibroblast growth factor 21 gene expression. *BioMed Res Int.* (2017) 29:8402035. doi: 10.1155/2017/8402035

39. Gu M, Cosenza G, Iannaccone M, Macciotta NPP, Guo Y, Di Stasio L, et al. The single nucleotide polymorphism g. 133A>C in the stearoyl CoA desaturase gene (SCD) promoter affects gene expression and quali-quantitative properties of river buffalo milk. *J Dairy Sci.* (2019) 102:442–51. doi: 10.3168/jds.2018-15059

40. Li K, Liu Y, He X, Tao L, Jiang Y, Lan R, et al. novel SNP in the promoter region of IGF1 associated with Yunshang Black Goat kidding number via promoting transcription activity by SP1. *Front Cell Dev Biol.* (2022) 10:873095. doi: 10.3389/fcell.2022.873095

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 10 May 2023
DOI 10.3389/fvets.2023.1180797

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Zahoor Khan,
University of Agriculture, Dera Ismail Khan, Pakistan

REVIEWED BY
Biao Chen,
Jiangxi Agricultural University, China
Kerong Shi,
Shandong Agricultural University, China

*CORRESPONDENCE
Qingbin Luo
✉ qbluo@scau.edu.cn

RECEIVED 06 March 2023
ACCEPTED 24 April 2023
PUBLISHED 10 May 2023

CITATION
Ye M, Fan Z, Xu Y, Luan K, Guo L, Zhang S and Luo Q (2023) Exploring the association between fat-related traits in chickens and the RGS16 gene: insights from polymorphism and functional validation analysis.
Front. Vet. Sci. 10:1180797.
doi: 10.3389/fvets.2023.1180797

COPYRIGHT
© 2023 Ye, Fan, Xu, Luan, Guo, Zhang and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploring the association between fat-related traits in chickens and the RGS16 gene: insights from polymorphism and functional validation analysis

Mao Ye[1,2], Zhexia Fan[1,2], Yuhang Xu[1,2], Kang Luan[1,2], Lijin Guo[1,2], Siyu Zhang[1,2] and Qingbin Luo[1,2]*

[1]Department of Animal Genetics, Breeding and Reproduction, College of Animal Science, South China Agricultural University, Guangzhou, China, [2]Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, Key Laboratory of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, Guangzhou, China

Introduction: Excessive fat deposition in chickens can lead to reduced feed utilization and meat quality, resulting in significant economic losses for the broiler industry. Therefore, reducing fat deposition has become an important breeding objective in addition to achieving high broiler weight, growth rate, and feed conversion efficiency. In our previous studies, we observed high expression of Regulators of G Protein Signaling 16 Gene (*RGS16*) in high-fat individuals. This led us to speculate that *RGS16* might be involved in the process of fat deposition in chickens.

Methods: Thus, we conducted a polymorphism and functional analysis of the RGS16 gene to investigate its association with fat-related phenotypic traits in chickens. Using a mixed linear model (MLM), this study explored the relationship between RGS16 gene polymorphisms and fat-related traits for the first time. We identified 30 SNPs of *RGS16* in a population of Wens Sanhuang chickens, among which 8 SNPs were significantly associated with fat-related traits, including sebum thickness (ST), abdominal fat weight (AFW), and abdominal fat weight (AFR). Furthermore, our findings demonstrated that AFW, AFR, and ST showed significant associations with at least two or more out of the eight identified SNPs of RGS16. We also validated the role of *RGS16* in ICP-1 cells through various experimental methods, including RT-qPCR, CCK- 8, EdU assays, and oil red O staining.

Results: Our functional validation experiments showed that *RGS16* was highly expressed in the abdominal adipose tissue of high-fat chickens and played a critical role in the regulation of fat deposition by promoting preadipocyte differentiation and inhibiting their proliferation. Taken together, our findings suggest that *RGS16* polymorphisms are associated with fat-related traits in chickens. Moreover, the ectopic expression of *RGS16* could inhibit preadipocyte proliferation but promote preadipocyte differentiation.

Discussion: Based on our current findings, we propose that the RGS16 gene could serve as a powerful genetic marker for marker-assisted breeding of chicken fat-related traits.

KEYWORDS

***RGS16*, chicken, SNP, abdominal fat, association**

1. Introduction

Being overweight and obese is a risk factor for major noncommunicable diseases (NCDs), including cardiovascular disease, type 2 diabetes, and cancer (1, 2). China has made many efforts to combat obesity, including implementing national policies and programs to promote healthy life-styles and prevent the development of NCDs, however, these efforts have been insufficient in controlling the rapid increase in overweight and obesity rates in the country (3). As a model animal, chickens are helpful in our study of abdominal fat deposition and may have certain guiding significance for the treatment of obesity-related diseases.

In recent decades, high-density genetic selection has greatly improved the weight, growth rate, and feed conversion efficiency of broilers (4). However, this has also resulted in excessive fat deposition, particularly in the abdomen. This phenomenon can lead to reduced feed utilization efficiency and lower meat quality, thus resulting in wasted resources and environmental pollution, which greatly reduce the desire for consumption and the economic efficiency of producers. Therefore, reducing fat deposition is a key issue to be addressed in broiler production. At the same time, research on fat deposition in chickens has important scientific significance for treating obesity-related diseases, reducing feed waste, and improving economic efficiency (5).

The protein encoded by the Regulators of G Protein Signaling 16 Gene (*RGS16*) belongs to the "regulator of G protein signaling" family. G protein signaling is activated through the binding of extracellular ligands to G protein-coupled receptors (GPCRs) and inhibited inside cells by regulator of G protein signaling (RGS) proteins (6). The GPCR pathway has been shown to influence the metabolism of glucose and fatty acids and the onset of obesity and diabetes (7, 8). RGS proteins are GTPase-activating proteins (GAPs) of alpha subunits that control the intensity and duration of GPCR signaling. The results of studies in recent years have shown that the triglyceride content is significantly higher in the liver of transgenic *RGS16* mice than of nontrans genic mice, indicating that *RGS16* inhibits fatty acid oxidation in the liver (9). It was also found that *RGS16* overexpression promotes lipid droplet formation in rat 832/13 cells and affects the expression of key genes for enzymes that mediate lipid droplet formation (10). Single nucleotide polymorphisms (SNPs) have begun to be used in animal breeding research as highly stable molecular markers that can provide a wealth of information (11, 12).

To date, *RGS16* has been relatively little studied in chicken compared with the mammalian counterpart. In our previous work, to explore the differences in fat deposition by high and low-fat broilers (13), we sequenced the transcriptomes of the abdominal fat of Wens Sanhuang chickens and found that the expression of *RGS16* in abdominal fat was higher in high-fat individuals. Therefore, we hypothesized that *RGS16* might be involved in the process of fat deposition in chicken. We provide some theoretical basis and direction for the selection of low-fat broilers by verifying the role of *RGS16* in ICP-1 cell and analyzing the relationship between polymorphisms and fat-related traits including abdominal fat weight, abdominal fat rate and sebum thickness.

2. Materials and methods

2.1. Experimental animals and fat-related traits data

The F_2 population of 100-day-old Sanhuang chickens ($n = 439$) used for the slaughter experiments in this study were reared under the same environmental and management conditions by Wens Food Group Co., Ltd. (Yunfu, China). The animal experiments in this study were approved by the Animal Care Committee of South China Agricultural University (permit number: SCAU#2017015, 13 September 2017) (14). The fat-related traits including AFW, AFR and ST were measured and calculated according to the Performance terminology and measurements for poultry (NY/T823-2020).

2.2. DNA extraction, PCR, and DNA sequencing

Blood samples were collected from veins under the wings of all Sanhuang chickens ($n = 439$), and DNA extraction was performed according to the instructions included in the DNA extraction kit (OMEGA, Georgia, United States). All DNA samples were used to amplify the dsDNA of the *RGS16* 5′UTR, 3′UTR and CDS with 2× Taq MasterMix (CWBIO, Nanjing, China). The PCR reaction conditions were as follows: pre-denaturation at 94°C for 2 min, amplification at 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, repeated for 30 cycles, final amplification at 72°C for 2 min, and hold at 4°C indefinitely. Finally, the PCR products were purified and sequenced by Sangon Biotech (Shanghai, China).

2.3. Cell culture, treatment, and transfection

The immortalized chicken preadipocyte 1 (ICP-1) cells used in this study were provided by Li Hui's research group from Northeast Agricultural University (Heilongjiang, China). The basal medium used was DMEM/F12 (Gibco, United States) supplemented with 15% fetal bovine serum (FBS) (Gibco, United States) and 1% streptomycin/penicillin solution (Gibco, United States), the cells were cultured at 37°C and 5% CO_2 in a calculator (15).

2.4. Overexpression plasmid construction and siRNA synthesis

To construct a chicken *RGS16* overexpression plasmid, the complete CDS region of *RGS16* was cloned into the EcoRI and BamHI sites of pcDNA3.1 (Promega, New York, United States), and the plasmids were extracted according to the instructions included with the HiPure Plasmid/BAC EF Mini Kit (Magen, Guangzhou, China). The knockdown of *RGS16* (5′-GGACCATTGATGGCCATAA-3′) was performed using specific siRNA designed and synthesized by RiboBio Co., Ltd. (Guangzhou, China).

TABLE 1 Summary statistics of fat-related traits in the Wens Sanhuang chicken population.

Trait	N	Mean	SD	Min.	Max.	Var	C.V (%)
AFW (g)	439	98.43	29.57	28.40	193.00	874.13	30.04
AFR (%)	439	7.72	1.76	2.79	13.68	0.03	22.80
ST (mm)	439	7.53	1.64	1.16	12.83	2.70	21.78

N, number; SD, standard deviation; CV, coefficient of variation; AFW, abdominal fat weight; AFR, abdominal fat rate; ST, sebum thickness.

2.5. RNA extraction, cDNA synthesis, and quantitative real-time PCR

Total RNA was extracted from tissues or cells using Magzol Reagent (Magen, Guangzhou, China) following the manufacturer's instructions. PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa, Otsu, Japan) was used for cDNA preparation from total RNA. RT-qPCR was performed using a ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) in a Bio-Rad CFX96 Real-Time Detection System. Data were analyzed using the $2^{-\triangle\triangle Ct}$ method with chicken *GAPDH* as the reference gene (16). The RT-qPCR primers were designed in Primer-BLAST[1] and their detail were listed in Supplementary Table 1.

2.6. Oil red O staining

For oil red O staining, the ICP-1 cells were washed with PBS (Gbico, New York, United States) and then fixed with 4% paraformaldehyde solution (Sangon Biotech, Shanghai, China) for 30 min at room temperature. After fixation, the cells were rinsed twice with PBS, and oil red O staining solution (Sangon Biotech, Shanghai, China) was added followed by the incubation of cells for 60 min at room temperature. Then, the staining solution was removed, and the cells were washed three times with PBS. Image analysis was carried out using a DMi8 microscope (Leica, Wetzlar, Germany). The oil red O dye in stained cells was extracted with isopropanol solution and the absorption value was measured at 510 nm using a microreader (Bio-Tek, Vermont, United States).

2.7. 5-ethynyl-2′-deoxyuridine assay

EdU staining was performed using a Cell-Light EdU Apollo 488 *In Vitro* Kit (RiboBio, Guangzhou, China). For the EdU assay, ICP-1 cells were treated with EdU medium (1:1,000; RiboBio, Guangzhou, China) for 2 h at 37°C and then fixed in 4% paraformaldehyde solution (Yike, Guangzhou, China) for 30 min. The cells were subsequently permeabilized using 0.1% Triton X-100 solution (Gbico, New York, United States). The cells were incubated with the staining solution for 30 min at room temperature in the dark. Finally, the stained cells were scraped off with a cell scraper, collected in a 1.5 mL centrifuge tube and resuspended in 1 mL PBS. A BD Accouri C6 flow cytometer (BD Biosciences, San Jose, CA, United States) was utilized for the analysis of stained cells (16).

2.8. Cell counting kit-8 assay

For the CCK-8 assay, ICP-1 cells were inoculated in 96-well plates and transfected with plasmids or siRNA. Cell proliferation and viability were then monitored at 12, 24, 36, and 48 h using the CCK-8 kit (TransGen Biotech, Beijing, China) following the manufacturer's instructions. This analysis was performed using a microreader (Bio-Rad, Hercules, CA, United States) to measure the absorption value at 450 nm.

2.9. Statistical analysis

The MLM package in IBM SPSS software (version 26, IBM: International Business Machines Corporation) was used to analyze the association of gene polymorphisms and haplotypes with chicken fat-related traits. All results were represented as mean ± SEM. The MLM model is as follows:

$$Y = \mu + G + S + e$$

Where Y represents the observed value, μ represents the mean, G represents the fixed effect of genotypes, S represents the fixed effect of sex, and e is the random error. In addition, we used Haploview v.4.2 software for linkage disequilibrium analysis (17, 18).

Statistical analysis and plots generation were conducted using GraphPad Prism v.9.4 (CA, United States) and R studio v.4.2.1 software (MA, United States). Correlation coefficients were calculated using Pearson correlation analysis. An unpaired *t*-test (two-tailed) was used to assess the statistical significance between two groups. In multiple comparisons, significant differences were assessed using the LDS method.

3. Results

3.1. Phenotypic data

The descriptive statistics of fat-related traits in the F_2 population of Wens Sanhuang chickens ($n = 439$) was listed in Table 1. Among of them, average AFW was 98.43 g, the max value was 193.00 while the min value was 28.40 indicating a particularly evident inter-individual differences and a wide range of variation, with a coefficient of variation exceeding 30%. Average AFR was 7.72%, with a min value of 2.79% and a max value of 13.68%, and a CV value of 22.8%. Its CV value was much lower than that of AFW, indicating a potential positive correlation between AFW and body weight. For ST, the average value was 7.53 mm, with a min value of 1.16 mm and a max value of 12.83 mm. The CV of ST was 21.87%, which was closed to that of AFR. In general, the CV of fat-related traits in this population was large, which could be used to validate the association analysis between nucleotide polymorphism and fat traits.

1 https://www.ncbi.nlm.nih.gov/tools/primer-blast/

3.2. SNPs discovery and genotypes

To screen for polymorphisms, the flanking and exon regions of *RGS16*, with a length of 3,053 bp, were, respectively, amplified (Figure 1A). A total of 30 SNPs were identified in this region from the F_2 population of Wens Sanhuang chickens, and these were mapped to the genome sequence (Version: GRCg6a) in the Ensembl Database for identification (Table 2). The sequencing files of the 439 samples were analyzed using SnapGene v4.1.8 software to identify their genotypes for every individuals. Among of the 30 SNPs, the 6 SNPs were located in exon region but they were all synonymous variant. The statistics of genotype frequencies and allele frequencies are shown in Table 3. The 30 identified SNPs could all be divided into three genotypes. A chi-square test confirmed that all identified SNPs were in accordance with Hardy–Weinberg equilibrium ($p > 0.05$). Pearson correlation coefficients were used to analyze the correlations between fat-related traits, including AFW, AFR, and ST. As shown in Figure 1B, AFW was significantly and positively correlated with AFR ($\rho = 0.934$, $p < 0.001$), while ST is significantly and positively correlated with AFW and AFR ($\rho = 0.290$ for AFW, $\rho = 0.234$ for AFR, $p < 0.001$).

3.3. Association between SNPs in *RGS16* and chicken fat-related traits

To perform association analysis of genotypes with fat-related traits in chickens, the MLM package in IBM SPSS software was utilized. The full results of this analysis were listed in Supplementary Table 2. The SNPs significantly associated with one or more fat-related traits were shown in Table 4. In terms of ST, the CC and TC genotypes of both rs735029742 and rs731455923 were found to be significantly higher than the TT genotype ($p < 0.05$). Additionally, the TT genotype of rs738408023 was significantly higher than the TC genotype ($p < 0.05$). For AFW and AFR, the TT and TC genotypes were significantly higher than CC genotypes for rs741662332 and rs315021359, respectively, while the CC genotypes were significantly higher than TT and TC genotypes for rs733023691 ($p < 0.05$). The CC and TC genotypes of rs735761612 were found to have higher AFR than the TT genotype ($p < 0.05$), and the AA and AG genotypes of rs16622145 were found to have higher AFR than the GG genotype ($p < 0.05$).

3.4. Linkage disequilibrium and haplotype analysis of the *RGS16*

To better understand the relationship between SNPs and fat-related traits, the eight SNPs significantly associated with fat-related traits were further analyzed for linkage disequilibrium (LD) using Haploview software (Figure 1C). The LD plot showed two haplotype blocks, with block 1 including rs735029742, rs738408023, and rs735761612, and block 2 including rs741662332, rs731455923, rs315021359, and rs16622145. All SNP pairs within each block had high LD values ranging from 0.93 to 1.00. Both blocks were analyzed using the MLM model in relation to three fat-related traits. Supplementary Table 3 showed that block 1 was not significantly associated with any traits, while block 2 was significantly associated with AFR and ST. Table 5 presented the traits and relevant data indicating a significant association with block 2 haplotypes. For AFR, both H3H3 and H4H4 genotypes were significantly lower than H1H1, H1H2, H1H3, H1H4, H2H3, and H4H3 ($p < 0.05$). As for ST, both H2H2 and H3H3 genotypes were significantly lower than H1H1, H1H4, and H4H2 ($p < 0.05$). In addition, the H2H2 genotype was significantly lower than H1H3 ($p < 0.05$). These associations demonstrated that *RGS16* may have a potential role in regulating chicken fat deposition.

3.5. *RGS16* inhibits preadipocyte proliferation

In our previous study, we found that *RGS16* was highly expressed in individuals with high level of abdominal fat (accession ID:

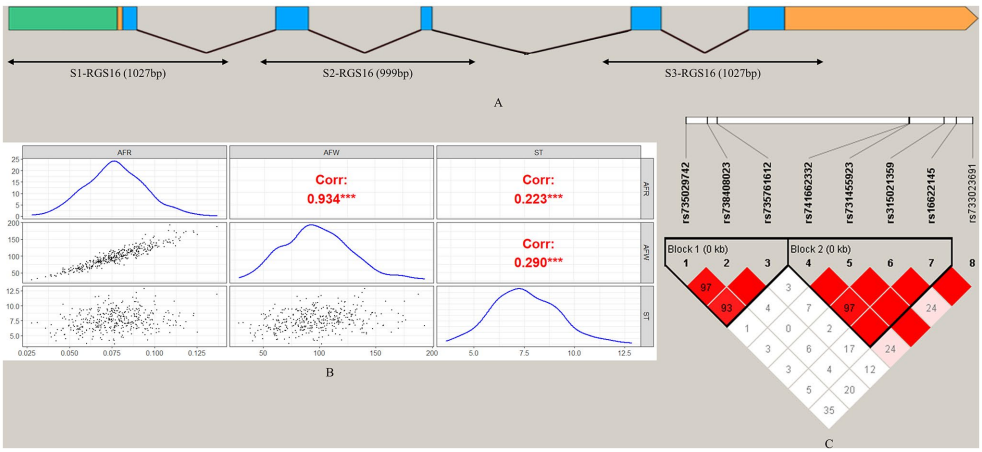


FIGURE 1
Analysis of 30 SNPs in the *RGS16* in Wens Sanhuang chickens. **(A)** The locations of the three primers used for SNP screening in the *RGS16*. **(B)** Pearson correlation coefficients between fat-related traits in Wens Sanhuang chickens. **(C)** The paired linkage disequilibrium (LD) values (D′) of the SNPs are represented by the values in the boxes. When D′=1, the values are not displayed. The intensity of the red color in the box represents the strength of LD, with darker shades indicating stronger LD. The Haploview software was used to define haplotype blocks with the default settings. * Denotes statistical significance with *<0.05; **<0.01; ***<0.001; ****<0.0001.

TABLE 2 Details of SNPs.

SNPs	Chr:bp	Alleles	Class	Conseq. type
rs736356137	8:5981044	T/C	SNP	Upstream gene variant
rs740542560	8:5981078	A/G	SNP	Upstream gene variant
rs731768083	8:5981083	G/C	SNP	Upstream gene variant
rs737985212	8:5981113	C/T	SNP	Upstream gene variant
rs740327014	8:5981388	G/A	SNP	Upstream gene variant
rs741317696	8:5981575	C/T	SNP	5 prime UTR variant
rs731255195	8:5981677	G/A	SNP	Intron variant
rs738816885	8:5981722	G/A	SNP	Intron variant
rs739467979	8:5981728	T/C	SNP	Intron variant
rs731997801	8:5981743	A/G	SNP	Intron variant
rs741182381	8:5981751	T/C	SNP	Intron variant
rs735029742	8:5983002	T/C	SNP	Intron variant
rs738408023	8:5983141	T/C	SNP	Intron variant
rs735761612	8:5983206	C/T	SNP	Intron variant
rs735218667	8:5983393	G/A	SNP	Intron variant
rs317340874	8:5983420	G/C	SNP	Intron variant
rs740889485	8:5983563	T/C	SNP	Intron variant
rs741662332	8:5984436	T/C	SNP	Synonymous variant
rs731455923	8:5984445	T/C	SNP	Synonymous variant
rs741115314	8:5984535	C/T	SNP	Synonymous variant
rs734517162	8:5984595	G/A	SNP	Intron variant
rs315021359	8:5984665	T/C	SNP	Intron variant
rs16622145	8:5984743	A/G	SNP	Intron variant
rs16622146	8:5984747	C/T	SNP	Intron variant
rs16622147	8:5984756	C/T	SNP	Intron variant
rs16622148	8:5984798	T/C	SNP	Intron variant
rs733023691	8:5984846	T/C	SNP	Intron variant
rs312359940	8:5984919	A/T	SNP	Intron variant
rs734891921	8:5984945	C/T	SNP	Intron variant
rs80763227	8:5985125	C/T	SNP	Synonymous variant

PRJNA656618). Here, the *RGS16* mRNA level difference between high- and low-fat groups was verified using RT-qPCR (Figure 2A). To investigate the function of *RGS16* in preadipocytes, ICP-1 cells were transfected with the *RGS16* overexpression plasmid and siRNA. The RT-qPCR results showed that the expression of *RGS16* mRNA could be upregulated by more than 400 times and knocked down by about 40%, respectively (Figures 2B,C). CCK-8 assay was performed to measure the proliferation viability of ICP-1 cells after 12-, 24-, 36-, and 48-h transfection. From these data, it is evident that *RGS16* overexpression significantly inhibits cell proliferation at 24 and 48 h, whereas *RGS16* knockdown significantly promotes cell proliferation at 36 h (Figures 2D,E). To further assess the function of *RGS16*, the cell cycle phase was detected using flow cytometry after 48-h transfection. The results showed that the overexpression of *RGS16* prolongs the G1 phase, preventing cells from entering the S phase and inhibiting their proliferation (Figures 2F,G). Furthermore, we also detected the mRNA levels of several cell-cycle-associated genes, *CCNE1*, *CCNA1*, and *CCND1* using RT-qPCR and found that the overexpression of *RGS16* significantly inhibits the expression of *CCNA1*, *CCNE1*, and *CCND1*, while *RGS16* knockdown had the opposite effect, revealing its inhibitory effect on cell proliferation (Figures 2H,I). We performed an EdU assay to verify this conclusion. Forty-eight h after transfection, EdU staining was detected using flow cytometry, and the results showed that the proportion of EdU-stained cells significantly decreases with *RGS16* overexpression and increases with *RGS16* knockdown (Figures 2J–M). All of the above experimental results indicated that *RGS16* is able to inhibit preadipocyte proliferation.

3.6. *RGS16* promotes preadipocyte differentiation

In addition, we hypothesized that *RGS16* may be involved in regulating the process of chicken preadipocyte differentiation, based

TABLE 3 Genotypes and allele frequencies and diversity parameters of SNPs in the *RGS16*.

SNP	Genotype frequencies (n)			Allelic frequencies		*p*-value	Genetic polymorphism			
	AA	BB	AB	A	B		*PIC*	*Ho*	*Ne*	*He*
rs736356137	65 (0.159)	154 (0.377)	190 (0.465)	0.391	0.609	0.882	0.363	0.476	1.910	0.465
rs740542560	284 (0.694)	16 (0.039)	109 (0.267)	0.828	0.172	0.411	0.245	0.285	1.399	0.267
rs731768083	284 (0.694)	16 (0.039)	109 (0.267)	0.828	0.172	0.411	0.245	0.285	1.399	0.267
rs737985212	7 (0.017)	291 (0.711)	111 (0.271)	0.153	0.847	0.622	0.225	0.259	1.349	0.271
rs740327014	329 (0.804)	6 (0.015)	74 (0.181)	0.895	0.105	0.739	0.170	0.188	1.232	0.181
rs741317696	7 (0.017)	273 (0.667)	129 (0.315)	0.175	0.825	0.169	0.247	0.289	1.406	0.315
rs731255195	254 (0.621)	10 (0.024)	145 (0.355)	0.798	0.202	0.125	0.270	0.322	1.475	0.355
rs738816885	257 (0.628)	10 (0.024)	142 (0.347)	0.802	0.198	0.171	0.267	0.318	1.466	0.347
rs739467979	10 (0.024)	257 (0.628)	142 (0.347)	0.198	0.802	0.171	0.267	0.318	1.466	0.347
rs731997801	255 (0.623)	11 (0.027)	143 (0.350)	0.798	0.202	0.223	0.270	0.322	1.475	0.350
rs741182381	10 (0.024)	256 (0.626)	143 (0.350)	0.199	0.801	0.154	0.268	0.319	1.469	0.350
rs735029742	4 (0.010)	323 (0.788)	83 (0.202)	0.111	0.889	0.871	0.178	0.197	1.246	0.202
rs738408023	311 (0.759)	5 (0.012)	94 (0.229)	0.873	0.127	0.776	0.197	0.221	1.285	0.229
rs735761612	49 (0.120)	166 (0.405)	195 (0.476)	0.357	0.643	0.772	0.354	0.459	1.849	0.476
rs735218667	10 (0.024)	254 (0.620)	146 (0.356)	0.202	0.798	0.115	0.271	0.323	1.477	0.356
rs317340874	254 (0.62)	14 (0.034)	142 (0.346)	0.793	0.207	0.553	0.275	0.329	1.490	0.346
rs740889485	4 (0.010)	304 (0.766)	89 (0.224)	0.122	0.878	0.666	0.191	0.214	1.273	0.224
rs741662332	265 (0.604)	20 (0.046)	154 (0.351)	0.779	0.221	0.924	0.285	0.344	1.525	0.351
rs731455923	28 (0.064)	249 (0.567)	162 (0.369)	0.248	0.752	0.972	0.304	0.373	1.596	0.369
rs741115314	11 (0.025)	324 (0.738)	104 (0.237)	0.144	0.856	0.749	0.216	0.246	1.326	0.237
rs734517162	11 (0.025)	282 (0.642)	146 (0.333)	0.191	0.809	0.294	0.262	0.309	1.448	0.333
rs315021359	262 (0.597)	19 (0.043)	158 (0.360)	0.777	0.223	0.731	0.287	0.347	1.531	0.360
rs16622145	287 (0.654)	12 (0.027)	140 (0.319)	0.813	0.187	0.581	0.258	0.304	1.436	0.319
rs16622146	16 (0.036)	287 (0.654)	136 (0.310)	0.191	0.809	1.000	0.262	0.309	1.448	0.310
rs16622147	118 (0.269)	107 (0.244)	214 (0.487)	0.513	0.487	0.877	0.375	0.500	1.999	0.487
rs16622148	288 (0.656)	15 (0.034)	136 (0.310)	0.811	0.189	0.977	0.260	0.307	1.442	0.310
rs733023691	239 (0.544)	22 (0.050)	178 (0.405)	0.747	0.253	0.309	0.306	0.378	1.607	0.405
rs312359940	206 (0.469)	45 (0.103)	188 (0.428)	0.683	0.317	0.977	0.339	0.433	1.763	0.428
rs734891921	43 (0.098)	173 (0.394)	223 (0.508)	0.352	0.648	0.059	0.352	0.456	1.839	0.508
rs80763227	101 (0.230)	126 (0.287)	212 (0.483)	0.472	0.528	0.810	0.374	0.498	1.994	0.483

SNP, single nucleotide polymorphism (numbers in parentheses indicate the number of individuals); PIC, polymorphism information content; Ho, observed heterozygosity; He, expected heterozygosity; Ne, effective number of alleles. The test for Hardy–Weinberg equilibrium with p-value > 0.05, indicating that the population is in Hardy–Weinberg equilibrium.

on its mRNA level difference between high- and low-fat individuals. ICP-1 differentiation was induced with 80 μM sodium oleate, and cells were collected at five different time points (12, 24, 36, 48, and 60 h) for RNA extraction. Interestingly, we found that *RGS16* expression was significantly increased during ICP-1 cell differentiation (Figure 3A). The cells were collected for RNA extraction after 48-h *RGS16* overexpression and *RGS16* knock down, and RT-qPCR was performed to detect the expression level of genes associated with preadipocyte differentiation (including *PPARγ*, *CEBPβ*, and *APOA1*). The RT-qPCR results showed that *RGS16* overexpression increases the expression of preadipocyte differentiation-related genes, whereas *RGS16* knockdown decreases their expression (Figures 3B,C). After 12 h of transfection, the cells were induced in 80 μM sodium oleate medium for 48 h and stained with oil red O. The results show that *RGS16* overexpression promotes lipid droplet formation, whereas *RGS16* knockdown inhibits lipid droplet formation (Figures 3D–G). Our results demonstrated that *RGS16* indeed is capable of driving preadipocyte differentiation and lipid droplet formation.

4. Discussion

In this study, we conducted an association analysis between *RGS16* polymorphisms and fat-related traits, and identified 8 SNPs that are significantly associated with fat-related traits. In addition, we also have concluded that *RGS16* plays an important role in regulating abdominal fat deposition with the mechanism by promoting the differentiation of ICP-1 cells.

TABLE 4 Association of eight SNPs in *RGS16* with fat-related traits in Wens Sanhuang chicken.

SNP	Trait	Mean±SEM			*p*-value
rs735029742		CC (322)	TC (83)	TT (5)	
	ST (mm)	7.63 ± 1.60^{A}	7.24 ± 1.68^{B}	6.42 ± 0.85^{AB}	0.043
rs738408023		CC (6)	TC (93)	TT (311)	
	ST (mm)	7.00 ± 1.64	7.18 ± 1.64^{A}	7.65 ± 1.60^{A}	0.034
rs735761612		CC (166)	TC (196)	TT (48)	
	AFR (%)	7.74 ± 1.83^{A}	7.74 ± 1.70^{B}	7.04 ± 1.53^{AB}	0.034
rs741662332		CC (20)	TC (154)	TT (265)	
	AFW (g)	83.1 ± 29.92^{AB}	100.31 ± 29.77^{A}	98.49 ± 29.2^{B}	0.049
	AFR (%)	6.80 ± 1.9^{AB}	7.88 ± 1.75^{A}	7.69 ± 1.74^{B}	0.034
rs731455923		CC (249)	TC (162)	TT (28)	
	ST (mm)	7.63 ± 1.61^{A}	7.49 ± 1.65^{B}	6.83 ± 1.76^{AB}	0.047
rs315021359		CC (19)	TC (158)	TT (262)	
	AFW (g)	83.25 ± 30.98^{AB}	99.95 ± 29.66^{A}	98.61 ± 29.21^{B}	0.043
	AFR (%)	6.84 ± 2.05^{AB}	7.85 ± 1.74^{A}	7.70 ± 1.74^{B}	0.045
rs16622145		AA (287)	AG (140)	GG (12)	
	AFR (%)	7.80 ± 1.80^{A}	7.66 ± 1.65^{B}	6.48 ± 1.71^{AB}	0.036
rs733023691		CC (22)	TC (178)	TT (239)	
	AFW (g)	114.96 ± 33.61^{AB}	96.65 ± 29.6^{B}	98.23 ± 28.8^{A}	0.023
	AFR (%)	8.60 ± 1.90^{AB}	7.60 ± 1.78^{B}	7.72 ± 1.72^{A}	0.042

The same letters in the same row indicate a significant difference ($p < 0.05$), whereas different or no letters in the same row indicate no significant difference ($p > 0.05$).

TABLE 5 Association analysis of blocks with fat-related traits in Wens Sanhuang chicken.

LD block	SNP	Haplotype	Diplotype (n)	Fat-related trait	
				AFR (%)	ST (mm)
Block 2			H1H1 (52)	7.85 ± 1.59^{D}	7.91 ± 1.89^{B}
			H1H2 (76)	7.98 ± 1.89^{B}	7.35 ± 1.72
			H1H3 (59)	8.04 ± 1.66^{A}	7.59 ± 1.68^{D}
	rs741662332	H1: TCTA (0.339)	H1H4 (56)	7.74 ± 1.65^{E}	7.94 ± 1.59^{A}
	rs731455923	H2: TTTA (0.247)	H2H2 (28)	7.63 ± 1.72	6.83 ± 1.76^{ABCD}
	rs315021359	H3: CCCA (0.216)	H2H3 (48)	7.68 ± 1.94^{F}	7.50 ± 1.68
	rs16622145	H4: TCTG (0.186)	H3H3 (17)	6.66 ± 1.90^{ABCDEF}	6.80 ± 0.95^{ABC}
			H4H2 (37)	7.35 ± 1.69	7.77 ± 1.46^{C}
			H4H3 (44)	7.87 ± 1.63^{C}	7.40 ± 1.34
			H4H4 (12)	6.48 ± 1.71^{ABCDEF}	7.37 ± 1.28
			p-value	0.0317	0.0496

The same letters on in the same column indicate a significant difference ($p < 0.05$), whereas different or no letters in the same column indicate no significant difference ($p > 0.05$).

With the continuous advancement of technology, molecular marker-assisted breeding has become one of the technologies that has attracted attention and application in the field of poultry production. This technology utilizes molecular marker techniques and principles of genetics to quickly and accurately screen poultry breeds with excellent genetic characteristics, avoiding the tedious breeding process and long feeding cycles involved in traditional breeding methods (11, 19). In this study, according to the Pearson correlation coefficient, we found a significant correlation between ST and AFW and AFR. This suggests that ST may be related to fat deposition, which is consistent with previous research findings (20). We conducted an analysis of the associations between *RGS16* gene polymorphisms and fat-related traits. A total of 30 SNPs were identified, out of which eight SNPs (rs735029742, rs738408023, rs735761612, rs741662332, rs731455923, rs315021359, rs16622145, and rs733023691) were found to be significantly associated with fat-related traits, such as AFW, AFR, and ST. Through association analysis, it was found that individuals carrying the *RGS16* with the TT genotype of rs735029742, the CC genotype of rs738408023, and the TT genotype of rs731455923 exhibited lower ST. Individuals with the TT genotype of rs735761612 and the GG genotype of rs16622145 showed lower

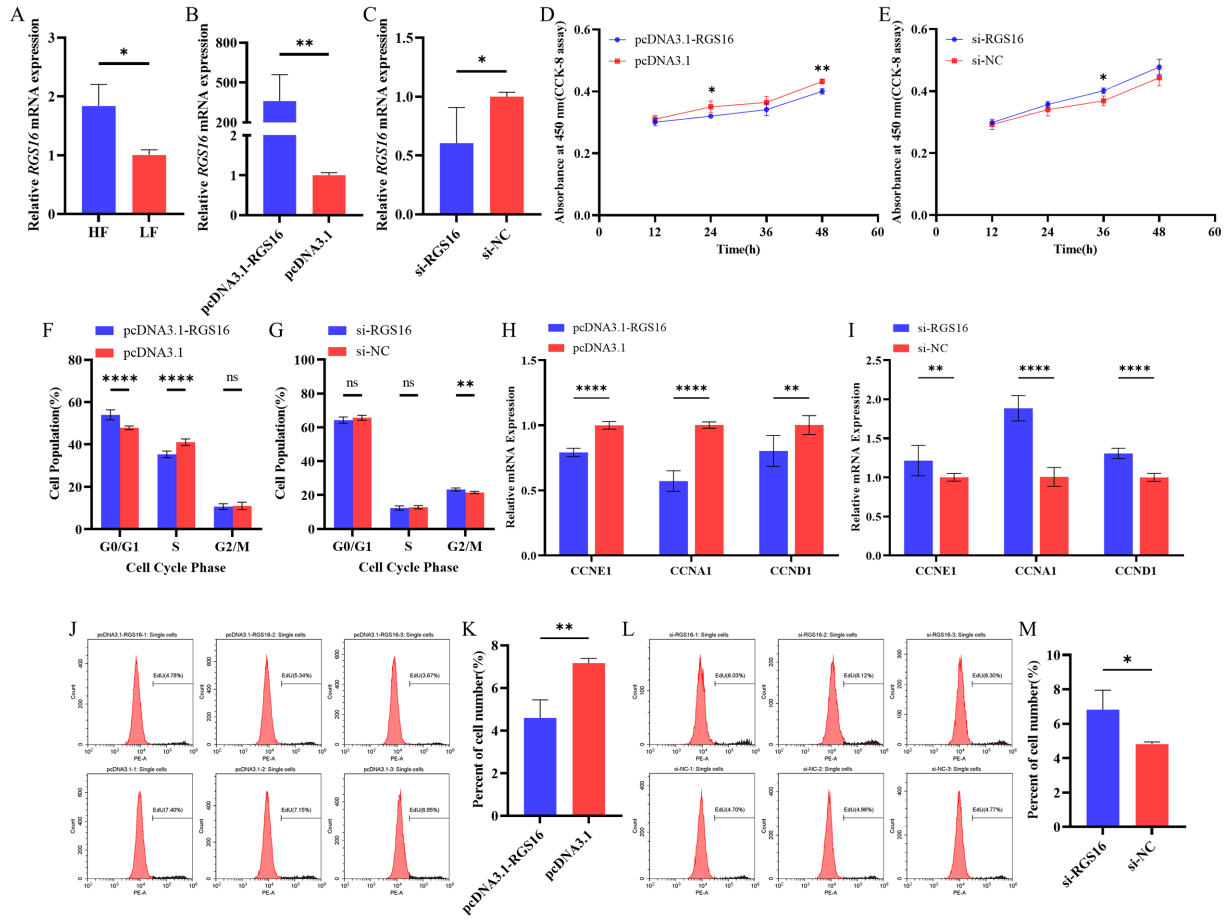


FIGURE 2
Regulation of preadipocyte proliferation by *RGS16*. **(A)** Differential expression of *RGS16* mRNA in the abdominal fat of high-fat and low-fat chickens measured by RT-qPCR (n=6 per group). **(B,C)** *RGS16* mRNA expression levels in ICP-1 cells overexpressing or with knockdown of *RGS16* (n=6 per group). **(D,E)** CCK-8 assay to assess the effect of *RGS16* overexpression or knockdown on ICP-1 cell viability (n=4 per group). **(F,G)** Cell cycle analysis of ICP-1 cells 48h after *RGS16* overexpression or knockdown (n=6 per group). **(H,I)** Expression of cell cycle genes in ICP-1 cells with *RGS16* overexpression and knockdown detected by RT-qPCR (n=6 per group). **(J–M)** ICP-1 cell cycle analysis using the flow cytometry EdU assay 48h after transfection with the *RGS16* overexpression vector or siRNA (n=3 per group). * Denotes statistical significance with *<0.05; **<0.01; ***<0.001; ****<0.0001.

AFR. Individuals with the CC genotype of rs741662332, the CC genotype of rs3152021359, and the TC genotype of rs733023691 showed lower AFR and AFW. With broiler breeders now focusing on selecting individuals with low fat weight or low rates of fat, these SNP genotypes could serve as molecular markers for improving fat-related traits in the Wens Sanhuang Chicken.

Synonymous mutations are associated with specific diseases or traits in multiple cases. When a cluster of infrequently used codons shifts from frequent codons to rare ones, it can affect the timing of co-translational folding and lead to changes in protein function (21–23). In this study, we identified four SNPs in the coding region of the *RGS16* gene, namely rs741662332, rs731455923, rs741115314, and rs80763227. However, these sites do not alter the amino acid sequence, which are also known as synonymous mutations. Although we could not directly observe the effects of synonymous mutations, we could infer that they may have an impact on these traits from the analysis of the effects of different genotypes of the synonymous SNP rs731455923 and rs315021359 on abdominal fat-related traits. In addition, we found that 6 SNPs in the intronic region of the RGS16 gene are significantly associated with adiposity-related traits. It should be noted that compared to SNPs in the coding region, the functional effects of SNPs in the intronic region were often more complex and difficult to predict. Specifically, intronic SNPs may affect the structure and function of RNA molecules, such as affecting splice site selection, regulating splicing efficiency, and influencing RNA stability and translation efficiency, which in turn affect protein expression and function (24–26). Haplotypes can often provide more information than a single SNP, as the phenotype of animals can be affected by multiple mutations (27). Subsequently, we used Haploview software to analyze the above eight SNPs for link-age disequilibrium. The finding that block 2 (rs741662332, rs731455923, rs315021359, and rs16622145) is related with AFR and ST offers strong evidence supporting the use of these SNPs as markers in breeding.

Adipogenesis and lipogenesis are regulated by a complex network of transcription factors that function at different stages of differentiation (28, 29). Peroxisome proliferator-activated receptor γ (*PPARγ*) and members of the CCAAT/enhancer-binding protein (*C/EBP*) family are key regulators of this process (13, 30, 31). As one of the three subtypes of the PPAR subfamily, *PPARγ* regulates the

10.3389/fvets.2023.1180797

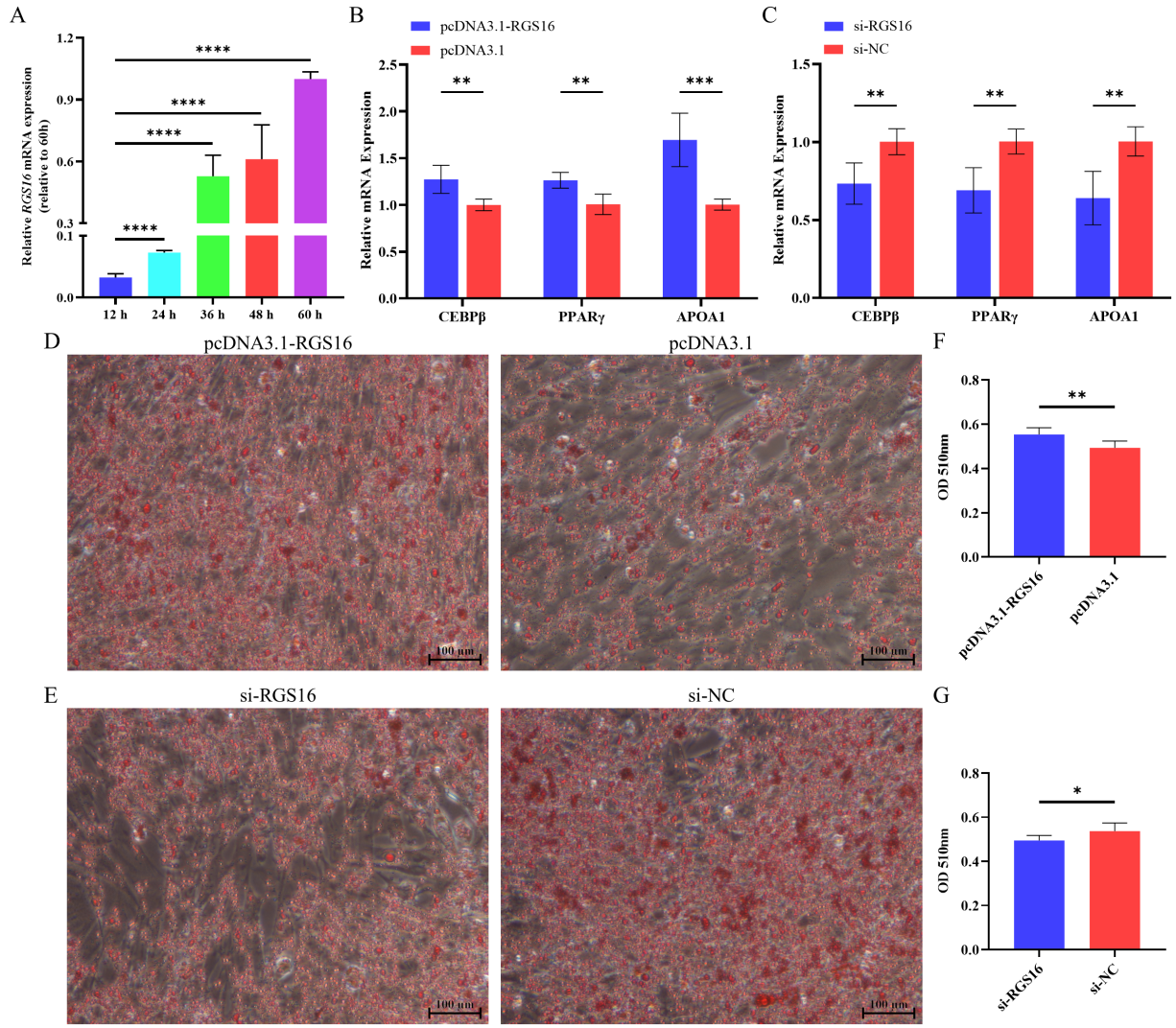


FIGURE 3
Regulation of preadipocyte differentiation by *RGS16*. **(A)** Expression of *RGS16* mRNA at different time points during differentiation measured by RT-qPCR (n=6 per group). **(B,C)** RT-qPCR analysis of preadipocyte differentiation-related gene expression in ICP-1 cells overexpressing or with knockdown of *RGS16* (n=6 per group). **(D–G)** Oil red O staining and quantification of cells on day 2 after transfection (n=6 per group). * Denotes statistical significance with *<0.05; **<0.01; ***<0.001; ****<0.0001.

lipogenesis of all adipocytes and binds to thousands of loci during the differentiation of white adipocytes (32, 33). It has been reported that *PPARγ* is highly expressed in high-fat chickens and that its expression increases during preadipocyte differentiation (34, 35). It has also been demonstrated that *C/EBPβ* activates the expression of *C/EBPα* and *PPARγ* through synergy with *C/EBPδ* to complete the adipocyte differentiation process (30, 36, 37). The characteristics of fat deposition differ between birds and mammals, where the main site of lipid biosynthesis is the liver or adipose tissue, respectively (38–40). *APOA1* is a component of high-density lipoprotein (HDL), a molecule that transports cholesterol and phospholipids from other parts of the body through the bloodstream to the liver (41, 42). In this study, we demonstrated that *RGS16* overexpression in ICP-1 cells significantly increases the expression of *PPARγ*, *APOA1*, and *C/EBPβ* and promotes lipid droplet formation. At the same time, we found that the expression of *RGS16* in ICP-1 cells in-creases with an increasing duration of differentiation. These results demonstrate that *RGS16* may promote the differentiation of preadipocytes in influencing adipogenesis.

5. Conclusion

In summary, *RGS16* regulates abdominal lipid deposition by promoting ICP-1 cell differentiation and inhibiting ICP-1 cell proliferation. Eight *RGS16* SNPs were found to be significantly associated with fat-related traits, including AFW, AFR, and ST, in the Wens Sanhuang chicken population. Of the identified SNPs, rs73145592 and rs315021359 were situated within coding region, while the remaining six polymorphisms (rs735029742, rs738408023, rs735761612, rs741662332, rs16622145, and rs733023691) were located within the intronic region of RGS16. Furthermore, the haplotypes within the LD block were found to be significantly associated with AFR and ST. Finally, the findings our current study indicate that *RGS16* may

play a significant role in the mechanism of abdominal fat accumulation in chickens. Therefore, it could be considered as a potential molecular marker to aid in the selection of broilers during breeding to improve their traits related to fat deposition.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MY and ZF: conceptualization. MY, QL, and SZ: methodology. MY: software, formal analysis, resources, data curation, writing—original draft preparation, and visualization. MY, KL, and YX: validation. ZF: investigation. LG and QL: writing—review and editing. QL: supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript..

Funding

Funding for this project was provided by several sources, including the Key-Area Research and Development Program of Guangdong Province (2022B0202100002), the Guangdong Province Modern Agricultural Industry Technology System Project (2022KJ128, 2023KJ128), the Science and Technology Program of Chaozhou City (202101ZD07), and the National Key Research and Development Program of China (2022YFF1000201).

Acknowledgments

We are very grateful to Li Hui from Northeast Agricultural University for providing us with ICP-1 cells.

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/fvets.2023.1180797/full#supplementary-material

References

1. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. (2008) 371:569–78. doi: 10.1016/S0140-6736(08)60269-X

2. Kivimäki M, Kuosma E, Ferrie JE, Luukkonen R, Nyberg ST, Alfredsson L, et al. Overweight, obesity, and risk of cardiometabolic multimorbidity: pooled analysis of individual-level data for 120 813 adults from 16 cohort studies from the USA and Europe. *Lancet Public Health*. (2017) 2:e277–85. doi: 10.1016/S2468-2667(17)30074-9

3. Collaborators GBDO, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med*. (2017) 377:13–27. doi: 10.1056/NEJMoa1614362

4. Huang HY, Liu RR, Zhao GP, Li QH, Zheng MQ, Zhang JJ, et al. Integrated analysis of microRNA and mRNA expression profiles in abdominal adipose tissues in chickens. *Sci Rep*. (2015) 5:16132. doi: 10.1038/srep16132

5. Nematbakhsh S, Pei Pei C, Selamat J, Nordin N, Idris LH, Abdull Razis AF. Molecular regulation of lipogenesis, adipogenesis and fat deposition in chicken. *Genes*. (2021) 12:414. doi: 10.3390/genes12030414

6. Huang J, Pashkov V, Kurrasch DM, Yu K, Gold SJ, Wilkie TM. Feeding and fasting controls liver expression of a regulator of G protein signaling (Rgs16) in periportal hepatocytes. *Comp Hepatol*. (2006) 5:8. doi: 10.1186/1476-5926-5-8

7. Alhosaini K, Azhar A, Alonazi A, Al-Zoghaibi F. Gpcrs: the Most promiscuous druggable receptor of the mankind. *Saudi Pharm J*. (2021) 29:539–51. doi: 10.1016/j.jsps.2021.04.015

8. Gendaszewska-Darmach E, Drzazga A, Koziolkiewicz M. Targeting GPCRs activated by fatty acid-derived lipids in type 2 diabetes. *Trends Mol Med*. (2019) 25:915–29. doi: 10.1016/j.molmed.2019.07.003

9. Pashkov V, Huang J, Parameswara VK, Kedzierski W, Kurrasch DM, Tall GG, et al. Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner. *J Biol Chem*. (2011) 286:15116–25. doi: 10.1074/jbc.M110.216234

10. Sae-Lee C, Moolsuwan K, Chan L, Poungvarin N. ChREBP regulates itself and metabolic genes implicated in lipid accumulation in beta-cell line. *PLoS One*. (2016) 11:e0147411. doi: 10.1371/journal.pone.0147411

11. Zhou Z, Cai D, Wei G, Cai B, Kong S, Ma M, et al. Polymorphisms of CRELD1 and DNAJC30 and their relationship with chicken carcass traits. *Poult Sci*. (2023) 102:102324. doi: 10.1016/j.psj.2022.102324

12. He S, Ren T, Lin W, Yang X, Hao T, Zhao G, et al. Identification of candidate genes associated with skin yellowness in yellow chickens. *Poult Sci*. (2023) 102:102469. doi: 10.1016/j.psj.2022.102469

13. Chao X, Guo L, Wang Q, Huang W, Liu M, Luan K, et al. miR-429-3p/LPIN1 Axis promotes chicken abdominal fat deposition via PPARγ pathway. *Front Cell Dev Biol*. (2020) 8:595637. doi: 10.3389/fcell.2020.595637

14. Guo LJ, Huang WL, Zhang SY, Huang YL, Xu YB, Wu RQ, et al. Chicken protein S gene regulates adipogenesis and affects abdominal fat deposition. *Animals*. (2022) 12:2046. doi: 10.3390/ani12162046

15. Li ZH, Zheng M, Mo JW, Li K, Yang X, Guo LJ, et al. Single-cell RNA sequencing of preadipocytes reveals the cell fate heterogeneity induced by melatonin. *J Pineal Res*. (2021) 70:e12725. doi: 10.1111/jpi.12725

16. Zhang J, Cai BL, Ma MT, Kong SF, Zhou Z, Zhang XQ, et al. LncRNA *SMARCD3-OT1* promotes muscle hypertrophy and fast-twitch fiber transformation via enhancing SMARCD3X4 expression. *Int J Mol Sci*. (2022) 23:4510. doi: 10.3390/ijms23094510

17. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. (2001) 68:978–89. doi: 10.1086/319501

18. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet*. (2005) 76:449–62. doi: 10.1086/428594

19. Cui HX, Shen QC, Zheng MQ, Su YC, Cai RC, Yu Y, et al. A selection method of chickens with blue-eggshell and dwarf traits by molecular marker-assisted selection. *Poult Sci*. (2019) 98:3114–8. doi: 10.3382/ps/pez069

20. Yang Z, Asare E, Yang Y, Yang JJ, Yang HM, Wang ZY. Dietary supplementation of betaine promotes lipolysis by regulating fatty acid metabolism in geese. *Poult Sci*. (2021) 100:101460. doi: 10.1016/j.psj.2021.101460

21. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*. (2007) 315:525–8. doi: 10.1126/science.1135308

22. Komar AA. Silent SNPs: impact on gene function and phenotype. *Pharmacogenomics*. (2007) 8:1075–80. doi: 10.2217/14622416.8.8.1075

23. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. (2011) 12:683–91. doi: 10.1038/nrg3051

24. Jo BS, Choi SS. Introns: the functional benefits of introns in genomes. *Genomics Inform*. (2015) 13:112–8. doi: 10.5808/GI.2015.13.4.112

25. Wang D, Guo Y, Wrighton SA, Cooke GE, Sadee W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J*. (2011) 11:274–86. doi: 10.1038/tpj.2010.28

26. Jacobsson JA, Schioth HB, Fredriksson R. The impact of Intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO. *Obes Rev*. (2012) 13:1096–109. doi: 10.1111/j.1467-789X.2012.01025.x

27. Liu NJ, Zhang K, Zhao HY. Haplotype-association analysis. *Adv Genet*. (2008) 60:335–405. doi: 10.1016/S0065-2660(07)00414-2

28. Guo L, Chao X, Huang W, Li Z, Luan K, Ye M, et al. Whole transcriptome analysis reveals a potential regulatory mechanism of LncRNA-FNIP2/miR-24-3p/FNIP2 axis in chicken adipogenesis. *Front Cell Dev Biol*. (2021) 9:653798. doi: 10.3389/fcell.2021.653798

29. Lefterova MI, Haakonsson AK, Lazar MA, Mandrup S. PPARγ and the global map of Adipogenesis and beyond. *Trends Endocrinol Metab*. (2014) 25:293–302. doi: 10.1016/j.tem.2014.04.001

30. Tanaka T, Yoshida N, Kishimoto T, Akira S. Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. *EMBO J*. (1997) 16:7432–43. doi: 10.1093/emboj/16.24.7432

31. Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, et al. Genome-wide profiling of Ppargamma: RXR and RNA polymerase ii occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev*. (2008) 22:2953–67. doi: 10.1101/gad.501108

32. Kawai M, Rosen CJ. Ppargamma: a circadian transcription factor in adipogenesis and osteogenesis. *Nat Rev Endocrinol*. (2010) 6:629–36. doi: 10.1038/nrendo.2010.155

33. Lefterova MI, Zhang Y, Steger DJ, Schupp M, Schug J, Cristancho A, et al. Ppargamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev*. (2008) 22:2941–52. doi: 10.1101/gad.1709008

34. Sun YN, Gao Y, Qiao SP, Wang SZ, Duan K, Wang YX, et al. Epigenetic DNA methylation in the promoters of peroxisome proliferator-activated receptor gamma in chicken lines divergently selected for fatness. *J Anim Sci*. (2014) 92:48–53. doi: 10.2527/jas.2013-6962

35. Zhang M, Li F, Ma XF, Li WT, Jiang RR, Han RL, et al. Identification of differentially expressed genes and pathways between intramuscular and abdominal fat-derived preadipocyte differentiation of chickens *in vitro*. *BMC Genomics*. (2019) 20:743. doi: 10.1186/s12864-019-6116-0

36. Tang QQ, Grønborg M, Huang HY, Kim JW, Otto TC, Pandey A, et al. Sequential phosphorylation of CCAAT enhancer-binding protein β by MAPK and glycogen synthase kinase 3β is required for adipogenesis. *Proc Natl Acad Sci U S A*. (2005) 102:9766–71. doi: 10.1073/pnas.0503891102

37. Tang QQ, Otto TC, Lane MD. Ccaat/enhancer-binding protein beta is required for mitotic clonal expansion during adipogenesis. *Proc Natl Acad Sci U S A*. (2003) 100:850–5. doi: 10.1073/pnas.0337434100

38. Agbu P, Carthew RW. microRNA-mediated regulation of glucose and lipid metabolism. *Nat Rev Mol Cell Biol*. (2021) 22:425–38. doi: 10.1038/s41580-021-00354-w

39. Han CC, Wang JW, Li L, Wang L, Zhang ZX. The role of LXR alpha in goose primary hepatocyte lipogenesis. *Mol Cell Biochem*. (2009) 322:37–42. doi: 10.1007/s11010-008-9937-8

40. Leveille GA, Romsos DR, Yeh Y, O'Hea EK. Lipid biosynthesis in the Chick. A consideration of site of synthesis, influence of diet and possible regulatory mechanisms. *Poult Sci*. (1975) 54:1075–93. doi: 10.3382/ps.0541075

41. Cochran BJ, Ong KL, Manandhar B, Rye KA. APOA1: a protein with multiple therapeutic functions. *Curr Atheroscler Rep*. (2021) 23:11. doi: 10.1007/s11883-021-00906-7

42. Wu CY, Wang YX, Gong PF, Wang LJ, Liu C, Chen C, et al. Promoter methylation regulates ApoA-I gene transcription in chicken abdominal adipose tissue. *J Agr Food Chem*. (2019) 67:4535–44. doi: 10.1021/acs.jafc.9b00007

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 05 June 2023
DOI 10.3389/fvets.2023.1196950

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Zahoor Khan,
University of Agriculture,
Dera Ismail Khan, Pakistan

REVIEWED BY
Qudrat Ullah,
University of Agriculture,
Dera Ismail Khan, Pakistan
Lei Liu,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Jie Yang
yangjie234@xju.edu.cn

RECEIVED 30 March 2023
ACCEPTED 19 May 2023
PUBLISHED 05 June 2023

CITATION
Yao H, Liang X, Dou Z, Zhao Z, Ma W, Hao Z, Yan H, Wang Y, Wu Z, Chen G and Yang J (2023) Transcriptome analysis to identify candidate genes related to mammary gland development of Bactrian camel (*Camelus bactrianus*). *Front. Vet. Sci.* 10:1196950. doi: 10.3389/fvets.2023.1196950

COPYRIGHT
© 2023 Yao, Liang, Dou, Zhao, Ma, Hao, Yan, Wang, Wu, Chen and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Transcriptome analysis to identify candidate genes related to mammary gland development of Bactrian camel (*Camelus bactrianus*)

Huaibing Yao[1,2], Xiaorui Liang[1,2], Zhihua Dou[1,2], Zhongkai Zhao[1,2], Wanpeng Ma[3], Zelin Hao[1,2], Hui Yan[1,2], Yuzhuo Wang[4], Zhuangyuan Wu[4], Gangliang Chen[2,5] and Jie Yang[1,2]*

[1]Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Ürümqi, China, [2]Xinjiang Camel Industry Engineering Technology Research Center, Ürümqi, China, [3]College of Veterinary Medicine, Xinjiang Agricultural University, Ürümqi, China, [4]Xinjiang Altai Regional Animal Husbandry Veterinary Station, Altay, China, [5]Bactrian Camel Academy of Xinjiang, Wangyuan Camel Milk Limited Company, Altay, China

Introduction: The demand for camel milk, which has unique therapeutic properties, is increasing. The mammary gland is the organ in mammals responsible for the production and quality of milk. However, few studies have investigated the genes or pathways related to mammary gland growth and development in Bactrian camels. This study aimed to compare the morphological changes in mammary gland tissue and transcriptome expression profiles between young and adult female Bactrian camels and to explore the potential candidate genes and signaling pathways related to mammary gland development.

Methods: Three 2 years-old female camels and three 5 years-old adult female camels were maintained in the same environment. The parenchyma of the mammary gland tissue was sampled from the camels using percutaneous needle biopsy. Morphological changes were observed using hematoxylin-eosin staining. High-throughput RNA sequencing was performed using the Illumina HiSeq platform to analyze changes in the transcriptome between young and adult camels. Functional enrichment, pathway enrichment, and protein–protein interaction networks were also analyzed. Gene expression was verified using quantitative real-time polymerase chain reaction (qRT-PCR).

Results: Histomorphological analysis showed that the mammary ducts and mammary epithelial cells in adult female camels were greatly developed and differentiated from those in young camels. Transcriptome analysis showed that 2,851 differentially expressed genes were obtained in the adult camel group compared to the young camel group, of which 1,420 were upregulated, 1,431 were downregulated, and 2,419 encoded proteins. Functional enrichment analysis revealed that the upregulated genes were significantly enriched for 24 pathways, including the Hedgehog signaling pathway which is closely related to mammary gland development. The downregulated genes were significantly enriched for seven pathways, among these the Wnt signaling pathway was significantly related to mammary gland development. The protein–protein interaction network sorted the nodes according to the degree of gene interaction and identified nine candidate genes: *PRKAB2*, *PRKAG3*, *PLCB4*, *BTRC*, *GLI1*, *WIF1*, *DKK2*, *FZD3*, and

WNT4. The expression of fifteen genes randomly detected by qRT-PCR showed results consistent with those of the transcriptome analysis.

Discussion: Preliminary findings indicate that the Hedgehog, Wnt, oxytocin, insulin, and steroid biosynthesis signaling pathways have important effects on mammary gland development in dairy camels. Given the importance of these pathways and the interconnections of the involved genes, the genes in these pathways should be considered potential candidate genes. This study provides a theoretical basis for elucidating the molecular mechanisms associated with mammary gland development and milk production in Bactrian camels.

KEYWORDS

Bactrian camel, transcriptome sequencing, mammary gland, development, differentially expressed gene

1. Introduction

Bactrian camels (two-humped camels), are mammals of the genus Camelidae of the family Mammalia that mainly inhabit cooler areas in Asia (1). There are five local species of Bactrian camels in China, namely Junggar, Tarim, Alashan, Sunit, and Qinghai, which are distributed in the arid and semi-arid regions of the Xinjiang, Inner Mongolia, Gansu, and Qinghai provinces in northwestern China. The Bactrian camel grazes a wide range of desert plants, can tolerate extreme environments, and has a specialized immune system. It is a distinctly important economic animal resource in desert areas, providing livestock products such as milk, meat, hides, and wool to herders in remote areas. In addition, camel milk has a high nutritional value, is hypoallergenic, and contains small-molecule nano-antibodies that can assist in the treatment of multifunctional disorders and autoimmune diseases such as lung cancer and diabetes (2, 3). However, camels grow slowly, reaching puberty at a later age than other livestock species. Sexual maturity is probably reached at 3–4 years. Female Bactrian camels give birth to their first calves at the age of 4 or 5 and produce offspring once every two years with a gestation period of 13–14 months. This slow reproductive rate, combined with the small geographical distribution, free-range grazing, and traditional breeding, results in low milk production and a shortage of camel milk supply to the market. Camels are crucial to the economies of many countries in arid and semi-arid regions of the world and camel breeding is thus increasing annually worldwide. With this increase, the metabolic processes and molecular mechanisms related to adaptation to extreme desert environments have been identified by genome sequencing and comparative genome analysis of Bactrian camels and dromedaries (4). There is a crucial need for developing knowledge of the genetic and molecular mechanisms related to milk production traits in Bactrian camels. Elucidating the mechanisms of mammary gland development is an initial key step to achieve this.

The mammary gland develops and produces milk under the regulation of systemic hormones, which is a dynamic and highly complex multistep process involving the periodic cycling of mammary epithelial proliferation, differentiation, and apoptosis (5). The components of milk are derived from blood-circulating nutrients, and they are synthesized in the epithelial cells of the mammary gland in lactating animals. Almost all studies on milk production traits have been conducted based on analysis of mammary glands and animal hematology (6). Mammary tissues undergo rapid growth during adolescence and develop again during pregnancy until lactation is complete (7). Animal studies have revealed apparent anatomical and physiological differences between juvenile and adult mammary glands. Therefore, the development and function of the mammary glands are essential for the provision of nutrition to offspring and the study of milk production traits. However, there are limitations in the analysis of milk performance traits using mammary tissues, such as difficulties in sampling, damage to animas, and ethical considerations.

Functional genomic tools have been widely used to study the molecular mechanisms of growth, development, and production traits in livestock (8, 9). Transcriptome analysis has been used to identify the molecular information that regulates economically important traits of an organism. Recently, researchers have studied the molecular mechanisms underlying mammary gland development in various species, such as cattle (10, 11), sheep (12), cats (13), rabbits (13), and rats (14). However, to date, the mechanisms underlying mammary gland development and the initiation of lactation have not been reported in Bactrian camels. The effects of specific candidate genes on physiology, mammary gland development and milk production traits remain unknown. Therefore, the current study uses transcriptome analysis of the mammary gland tissues of young and adult camels to investigate important candidate genes and pathways involved in mammary gland development in Bactrian camels.

2. Materials and methods

2.1. Experimental camels and sample collection

The transcriptome sequencing sample population consisted of 6 domestic Junggar Bactrian camels with no consanguineous relationships among them, including adolescent young camels (YTR; $n = 3$; 2 years old) and lactating adult female camels (CNR; $n = 3$; 5 years old) from Fuhai County, Altay Prefecture, northwest China's Xinjiang Uygur Autonomous Region (87° 35′25″E, 46° 86′6″N). The mammary glands of 2 years-old adolescent female camels are in the developmental stage and do not lactate. At 4 or 5 years of age, female camels begin to sexually mature until the end of their first pregnancy, when their mammary glands are fully developed and they have the

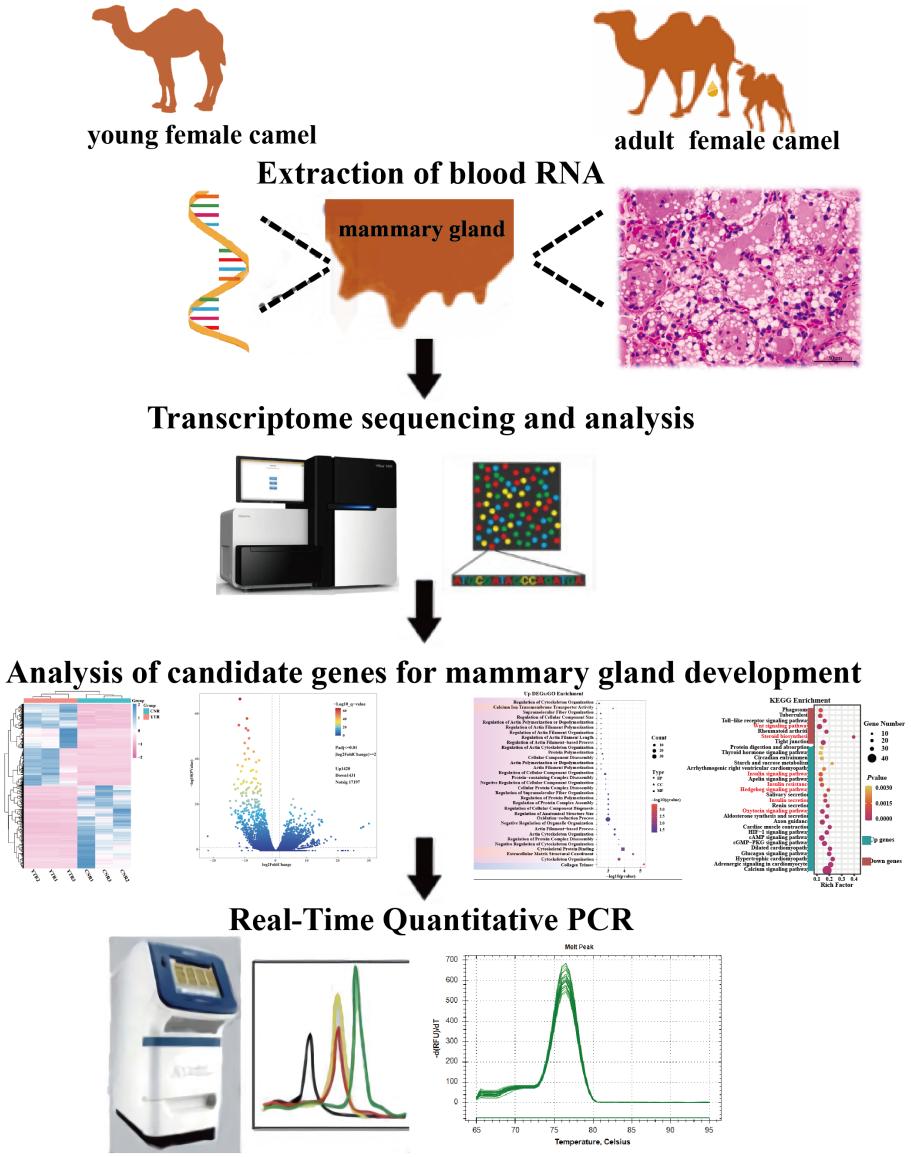


FIGURE 1
Experimental design of the study.

ability to secrete milk after giving birth. All experimental camels were fed under the same conditions to reduce the environmental effects on gene expression.

A specialist veterinarian conducted sampling to minimize harm to the camels. The parenchyma of the mammary gland tissue, which contains lobules that synthesize milk (15), was harvested from camels by percutaneous needle biopsy. Part of the tissue samples were immediately frozen in liquid nitrogen then transferred to the laboratory and stored at −80°C for subsequent RNA extraction. The remaining mammary gland tissues were washed with sterile saline and preserved in a 2.5% glutaraldehyde solution for later use in hematoxylin-eosin (H&E) staining (Figure 1). All the experimental procedures were performed in accordance with the guidelines of the Laboratory Animal Administration Regulations issued by the National Science and Technology Committee (China). All experimental animal studies were reviewed and approved by the Animal Welfare Committee of Xinjiang University (approval ID: XJU2019012).

2.2. Micromorphological examination

The mammary gland tissues obtained from the experimental camels were fixed in 10% neutral buffered formalin, dehydrated in a graded alcohol series, cleared in xylol, embedded in paraffin, and sectioned to a thickness of 4–5 µm. Tissue sections were prepared and stained with H&E following the protocols of a previous study (16) and photographed under a light microscope (Eclipse E100 Nikon, Japan).

2.3. RNA extraction, quality, and integrity determination

Total RNA was extracted from the camel mammary tissue using a TRIZOL RNA Extraction Kit (Invitrogen, Carlsbad, CA, United States). RNA integrity was evaluated using agarose 1.0% gel electrophoresis and the RNA Nano 6,000 Assay Kit of the Agilent 2,100 Bioanalyzer (Agilent, Santa Clara, CA, United States). RNA samples that passed quality control were used for subsequent sequencing library construction.

2.4. Library construction and transcriptome sequencing

Library preparation and transcriptome sequencing were performed by Novogene Co., Ltd. (Tianjin, China). The mRNA was purified from 1 μg of total RNA using oligo dT magnetic beads. mRNA was fragmented using divalent cations at high temperatures. After fragmentation, the cDNA library was synthesized using random hexamer primers, M-MuLV Reverse Transcriptase, and paired-end sequencing using the HiSeq 6000 platform (Illumina) (17). Reads containing adapters, poly-N, and low-quality reads were removed to obtain clean reads by Fastp (v 0.19.7). The Q20, Q30, and GC contents of the clean data were calculated and all downstream analyses were performed using high-quality clean data. A set of genomic index files of *Camelus bactrianus* reference genome was built using Hisat2 (version 2.0.5) and clean paired-end reads were mapped to the reference genome using Hisat2. Gene expression levels were quantified using featureCounts (version 1.5.0-p3) and the fragments per kilobase of transcript per million mapped reads (FPKM) of each annotated gene were counted based on the length of the gene and the read counts mapped to this gene (18).

2.5. Differentially expressed genes identification

Principal component analysis (PCA) was used to visually identify differences between the different sequencing sample groups in the transcriptome data. Differential expression analysis was performed between the two groups of camels using the DESeq2 package in R (version 4.2.1), based on negative binomial distribution. Considering previous studies (19, 20) and our specific experimental situation, adjusted *p*-values (p_{adjust}) < 0.01 and |log2FoldChange| >2 were established as the thresholds for significance. *p*-values were adjusted for multiple testing using the Benjamini and Hochberg methods. Genes with p_{adjust} < 0.01 and log2FoldChange >2 by DESeq2 were defined as upregulated differentially expressed genes (DEGs); genes with p_{adjust} < 0.01 and log2FoldChange <−2 were regarded as downregulated DEGs.

2.6. Gene functional enrichment analysis of DEGs

To explore functions and pathways associated with the DEGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the NovoMagic Cloud Platform.[1] A cut-off of $p < 0.05$ was used to screen significant functions and pathways. We further screened the core pathways and candidate genes by systematically reviewing the literature.

2.7. Protein–protein interaction network of DEGs

Gene regulatory networks have elucidated the regulation of mammary gland development in animals (21, 22). Therefore, we uploaded all identified DEGs to the Search Tool for the Retrieval of Interacting Genes/Proteins database[2] to create a protein–protein interaction (PPI) network (23). PPI networks were established using Cytoscape software (version 4.8.0). Hub genes were identified and screened by network analysis using Cytoscape and its plugin (CytoHubba) (24).

2.8. Quantitative real-time polymerase chain reaction validation of DEGs

The Prime Script RT Reagent Kit (Takara Biotechnology Co., Ltd., Shiga, Japan) was used to produce cDNA from RNA. The obtained cDNA was analyzed immediately or stored at −20°C. Camel beta-actin (β-actin) (NCBI accession no. XM_010965866.2) was used as an internal reference gene (25). Fifteen genes were selected from the DEGs between the young camel and adult camel groups at another sampling site. Polymerase chain reaction (PCR) was performed using the following cycling conditions: 95°C for 5 min; 40 cycles of 95°C for 10 s and 60°C for 30 s; 95°C for 15 s, 60°C for 60 s, and 95°C for 15 s. Specific primers were designed using the Primer-BLAST online website. Three replicates of each sample were used to calculate the relative expression by the $2^{-\Delta\Delta ct}$ method (26). The identities of these DEGs and specific primer sequences are listed in Table 1.

3. Results

3.1. Histomorphometric analysis of mammary gland

We first examined the morphological and microscopic characteristics of camel mammary glands according to published literature on other species. The morphological differences at different stages of mammary gland development are shown in Figure 2. As reported for other young animals (27–29), the mammary gland tissue of young camels was incompletely developed and consisted of a large amount of connective tissue (Figures 2A,B). Furthermore, the terminal duct lobular units (TDLU) in the mammary gland were incompletely developed, with inner epithelial tissue surrounded by lobular connective tissue and an outer overlay of interlobular

1 https://magic.novogene.com/customer/main#/loginNew
2 http://string-db.org/

10.3389/fvets.2023.1196950

TABLE 1 The primers used for RT-qPCR are listed (F=forward, R=reverse).

Gene	Primer sequence (5′→3′)	Annealing temperature (°C)	Product length (bp)
β-actin	F: CAGATCATGTTCGAGACCTTC	55	275
	R: ATGTCACGCACGATTTCC		
CAMK2D	F: CTCTCCTGTAGGAAGCAACCAG	56	101
	R: CCTTTGCCATCCATCCCACT		
CAMK2A	F: CGATGACTCTCCTTTTCTCC	54	105
	R: CTCCCAAGTTTCTTCTTGGA		
BTRC	F: GAGAGATACTATGGCTACACTG	55	114
	R: GCTCAGTGTTCTTTCCATCAG		
PRKAB2	F: GGATCCCGATGGAGAAGT	55	104
	R: GGAAGTCTGAAGTAGAAGCC		
PLCB4	F: CTGCTCACTCAAAAGGATCT	55	104
	R: GTGAGTGAAGTTTCTGGGTA		
PPP1R3A	F: TTAGGCACAGAAAGGTGAAG	56	102
	R: GCTCGTTTCACAGTTGAGTA		
SLC1A1	F: TACGCTTATTTCTGGGTGAG	54	102
	R: AACCTCCTTCTTCTCTTCCT		
LTF	F: ACCAAGGAAGAATCATCACC	55	111
	R: CCGGGTATTTGTTGTTAAGC		
PIGR	F: AGTTGCTCTCACCAATAAGG	55	109
	R: GTTCACTCTGCTTAGCTCTT		
GLI1	F: CAGAGAGACCAACAGCTGCA	54	112
	R: CAGGCTGGCATCCGATAGAG		
WIF1	F: CAGGCGAGAGTGCTCATAGG	56	118
	R: TGACAGGAATCGCTGGCATT		
DKK2	F: TCGGCACAGAGATCGAAACC	54	171
	R: TTGGTCCAGAAGTGACGAGC		
FZD3	F: CGATCGAGGAAGCATGGCTA	56	126
	R: TCTTGGCACATCCTCAAGGT		
WNT4	F: TATCCTGACACACATGCGGG	56	139
	R: CTCGGTGGCTCCATCAAACT		
KCNJ12	F: CATCACCATCCTGCACGAGA	54	160
	R: CTCATTGGCCAGGTAGGAGC		

connective tissue, and with fewer and irregularly shaped ducts and epithelial cells. The mammary glands of adult female camels were more fully developed (Figures 2C,D), with a large number of mammary epithelial cells and uniform chromatin distribution, adipose tissue, a small amount of thin connective tissue, and fat pads scattered with alveoli, ducts, and alveolar cavities.

3.2. RNA-sequencing reads and mapping to the reference genome

Raw reads were obtained for each library and between 43 and 59 million clean reads were obtained after quality control procedures and applied to the YTR and CNR libraries (Table 2). The Q30 reads (quality score >97.69%) were >95% for all samples (Supplementary Table S1). Between 38 and 53 million reads were mapped to the Bactrian camel genome for the YTR and CNR groups, respectively. Approximately 92% of the clean reads mapped to the camel reference genome, of which nearly 89% were uniquely mapped. The detailed statistical information is provided in Supplementary Table S2. These data provided a solid basis for subsequent analyses.

3.3. Analysis of DEGs

The number of genes at different expression intervals was determined (Supplementary Table S3). The first principal component

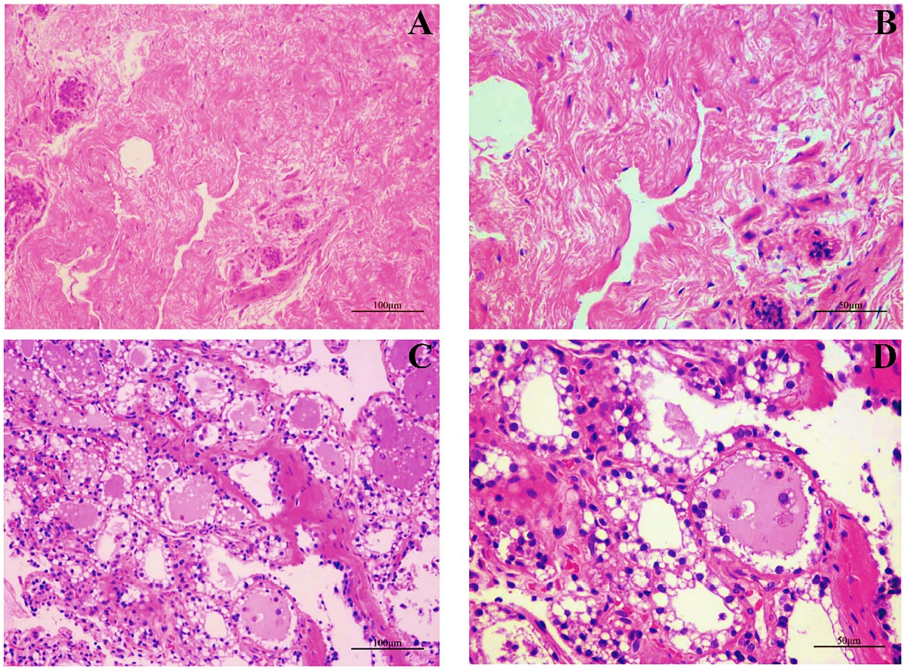


FIGURE 2
Mammary gland morphology in young and adult female camels. **(A)** Morphological structure of the mammary glands of young camels (200×). **(B)** Morphological structure of the mammary glands of young camels (400×). **(C)** Morphological structure of the mammary glands of adult camels (200×). **(D)** Morphological structure of the mammary glands of adult camels (400×).

TABLE 2 Summary of the RNA-seq analyses after mapping to the reference genome.

Sample	Raw reads	Clean reads	GC content	Total map	Unique map	Multi map
YTR1	59,165,478	58,350,742	51.57	53,783,775 (92.17%)	52,472,376 (89.93%)	1,311,399 (2.25%)
YTR2	46,255,466	45,642,826	51.48	42,138,839 (92.32%)	41,094,703 (90.04%)	1,044,136 (2.29%)
YTR3	43,870,868	41,787,324	48.54	38,482,571 (92.09%)	37,891,330 (90.68%)	591,241 (1.41%)
CNR1	46,672,290	45,317,128	42.60	42,437,154 (93.64%)	41,233,336 (90.99%)	1,203,818 (2.66%)
CNR2	47,309,962	46,649,412	47.39	44,449,779 (95.28%)	42,274,957 (90.62%)	2,174,822 (4.66%)
CNR3	49,864,878	49,116,410	45.86	47,175,819 (96.05%)	44,318,893 (90.23%)	2,856,926 (5.82%)

(PCA1), which had the largest variance (41.43%), distinctly clustered the samples into two groups, with three biological replicates within each group clustered together (Figure 3A). A total of 2,851 genes were identified as differentially expressed: 1,420 (49.81%) were upregulated and 1,431 (50.19%) were downregulated (YTR vs. CNR) (Figure 3B and Supplementary Table S4). The expressed genes were subjected to hierarchical clustering to comprehensively and intuitively display the differences in gene expression between the two groups (Figure 3C). Statistical analysis of the gene biotypes was performed on the DEGs to identify the composition of gene categories. Of these, 2,419 (84.8%) were protein-coding genes, followed by unknown type genes (10.8%) and 93 (3.3%) long non-coding RNA genes (Figure 3D).

3.4. Go enrichment analysis

To obtain a more comprehensive understanding of the DEGs, we performed GO functional enrichment analysis using DAVID. These DEGs were enriched in 827 GO terms, of which 434 were related to biological processes (BP), 279 to molecular functions (MF), and 114 to cellular components (CC). Moreover, collagen trimers and extracellular matrix structural constituents were significantly enriched ($p_{adjust} < 0.05$). The complete GO analysis results are provided in Supplementary Table S5. Significance analysis of GO enrichment of upregulated DEGs revealed 33 significantly enriched entries ($p_{adjust} < 0.05$), including 29 BP terms, 3 MF terms, and 1 CC term (Supplementary Table S6). These terms were related to cytoskeletal organization, organelle regulation, actin regulation, protein organization, cytoskeletal and component organization regulation, and calcium ion transport across membranes (Figure 4A).

GO enrichment of the downregulated DEGs revealed 13 significantly enriched entries ($p_{adjust} < 0.05$), including five BP terms, five MF terms, and three CC terms (Supplementary Table S7). These terms were involved in molecular functional regulators, riboprotein complexes, cytoplasmic fraction, peptidase regulator activity and inhibitor activity, ribosomes, translation, peptide metabolism and peptide biosynthesis processes, cellular amide metabolism processes, amide biosynthesis processes, ribosome structural components, and

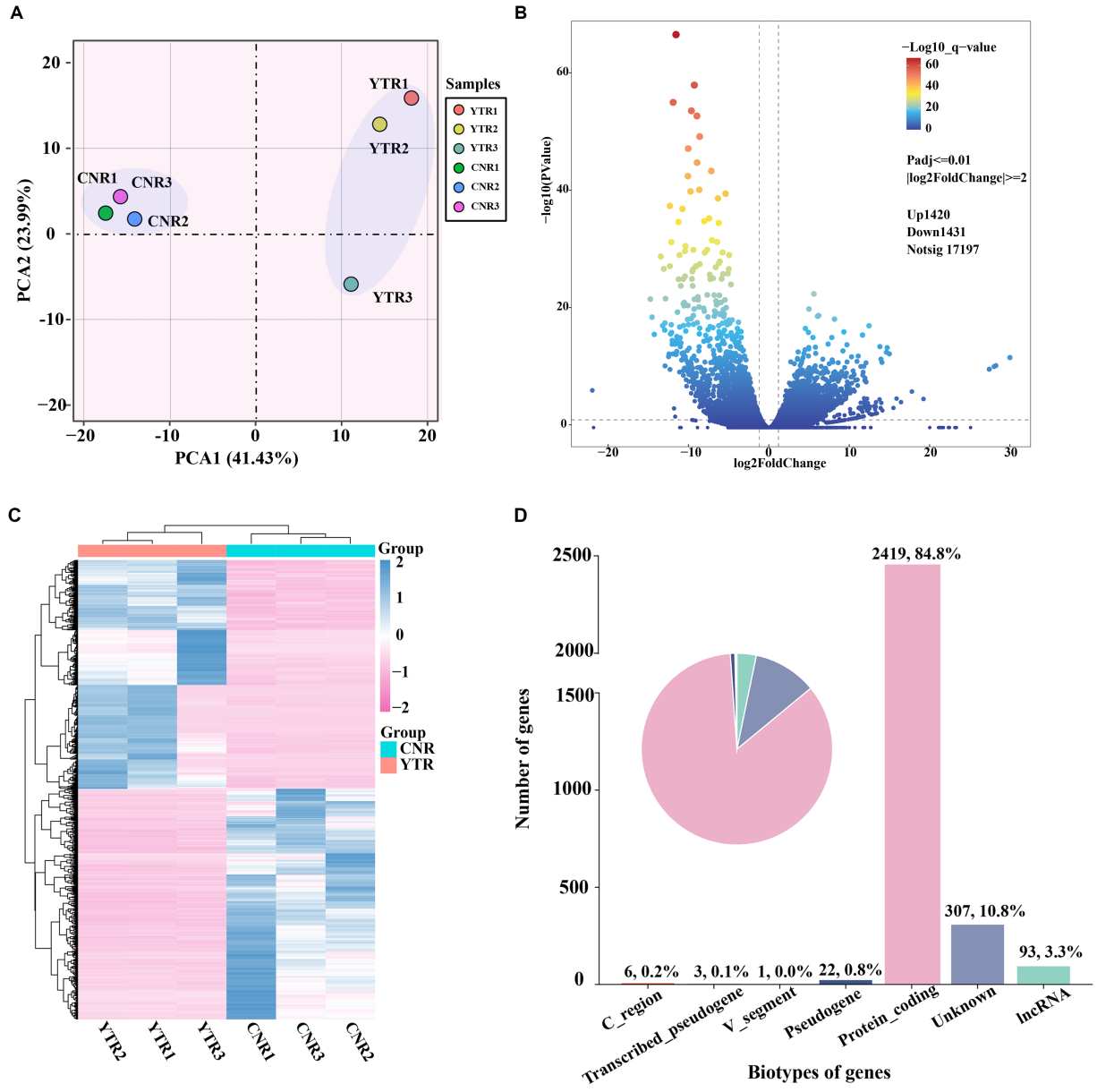


FIGURE 3
Analysis of differentially expressed genes among sample groups. **(A)** Principal-component analysis comparing the two sample groups. **(B)** Volcano plot of differentially expressed genes. **(C)** Heatmap and the hierarchical cluster analysis of the differentially expressed genes. **(D)** Biotype of genes statistical analysis.

structural molecular activity (Figure 4B). In the reference GO terms reported in previous studies, these significantly enriched GO entries were intimately correlated with the maintenance of cell morphology and structure, organismal tissue growth and development, and response to cell growth. Among these, peptidase regulator activity and inhibitor activity entries can regulate aminopeptidase N to promote mammary gland development in animals.

3.5. KEGG pathway enrichment analysis

Signaling pathways are essential for mammary gland development. We performed KEGG pathway enrichment analysis to investigate the gene-enriched pathways. Overall, 2,851 DEGs were assigned to 318 KEGG pathways, of which 11 were significantly enriched ($p_{adjust} < 0.05$). The main pathways were insulin secretion, calcium and glucagon signaling pathway, steroid biosynthesis, HIF-1 signaling pathway, protein digestion and absorption, adrenergic signaling in cardiomyocytes, and tight junction (Supplementary Table S8).

The upregulated and downregulated genes were subjected to KEGG pathway enrichment analysis. The enrichment analysis results indicated that 24 pathways were significantly up-regulated and 7 were significantly down-regulated ($p_{adjust} < 0.05$) (Figure 5). Upregulated genes were significantly enriched in the Hedgehog, insulin, and oxytocin signaling pathways (Supplementary Table S9). Downregulated genes were significantly enriched in the

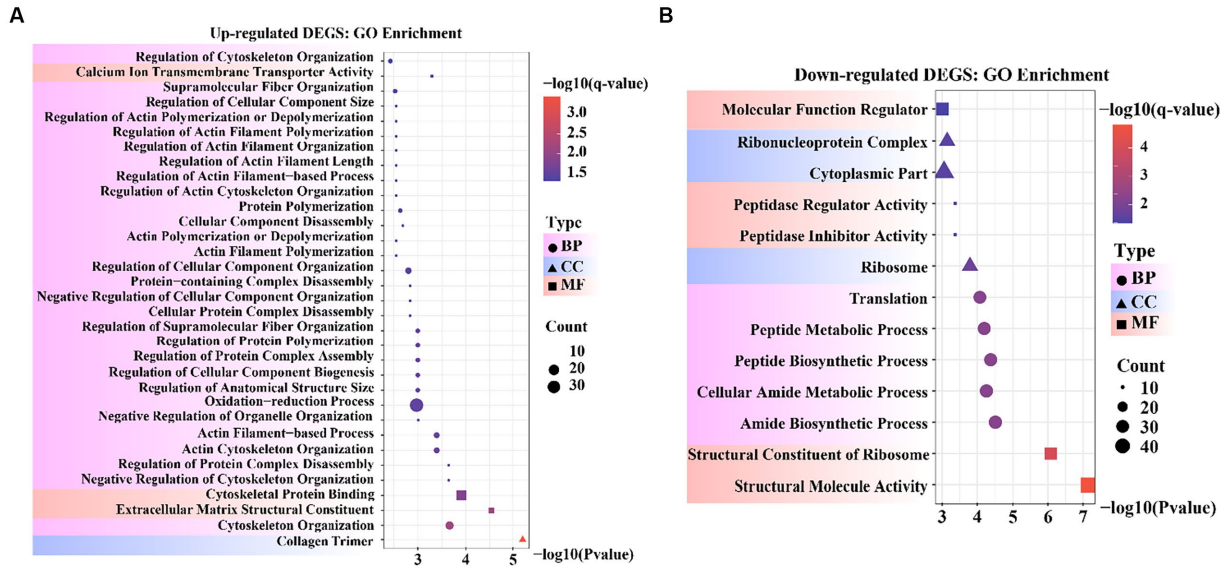


FIGURE 4
Bubble diagram of Gene Ontology (GO) enrichment result. **(A)** GO enrichment data for genes that were up-regulated. **(B)** GO enrichment data for genes that were down-regulated.

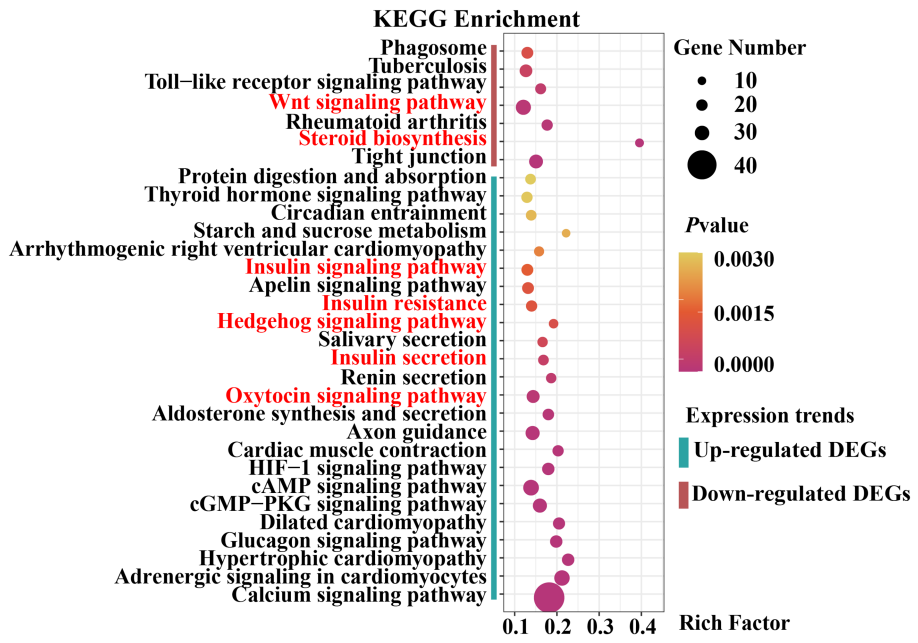


FIGURE 5
Dot plot of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis results.

Wingless-Type MMTV Integration Site Family (Wnt) signaling pathway, steroid biosynthesis, and tight junctions (Supplementary Table S10). As previously reported, the Hedgehog signaling pathway (30, 31), Wnt signaling pathway (32, 33), hormone-related pathways (34–36), and tight junctions (37) are closely associated with mammary gland development in animals. Collectively, these results provide a basis for selecting candidate genes related to mammary gland growth and development.

3.6. PPI network of the DEGs

We explored the biological and regulatory functions of hub DEGs at the protein level. Next, we collated the genes in important pathways for the (PPI) network analysis and sorted the nodes according to the degree of interaction (Figure 6 and Supplementary Table S11). Among the top 10 central genes, *PRKAB2* and *PRKAG3* are associated with animal growth, slaughter, and meat quality traits (38); *PLCB4* genes

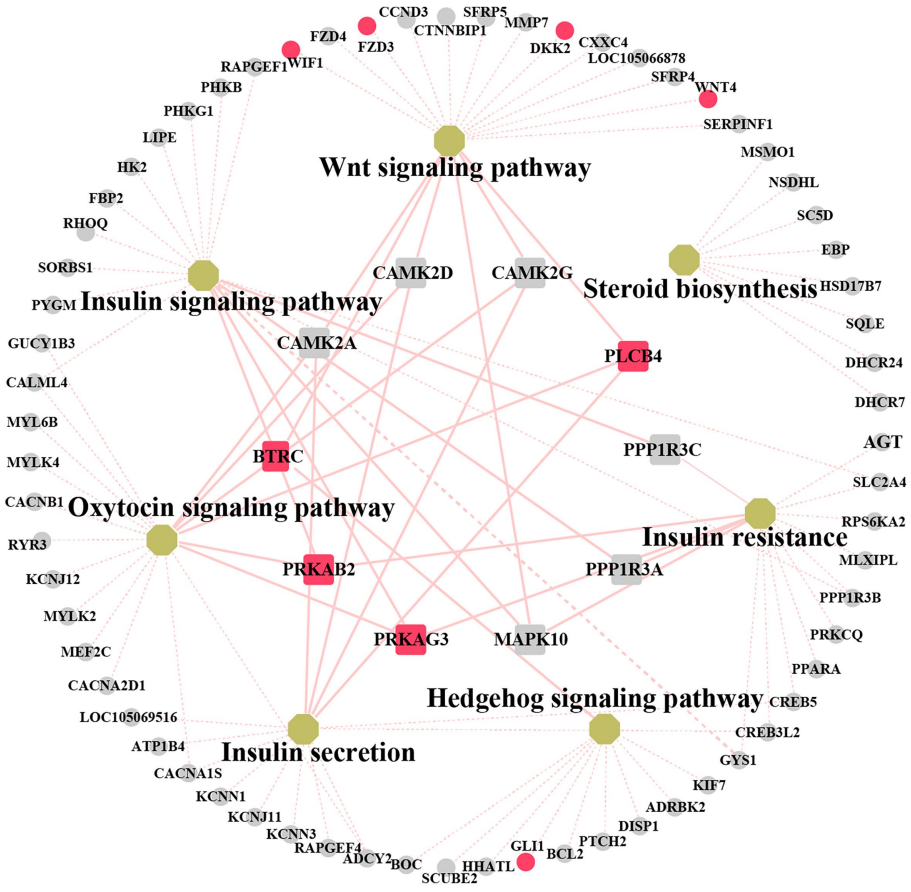


FIGURE 6
Networks analysis between protein–protein interaction (PPI) network and KEGG pathway related to mammary gland development. Non-hub genes are represented as grey circle while yellow octagons represent pathways. Hub genes are indicated by rectangles. The important candiate genes are highlighted in red. The light red solid line indicates the high degree of gene linkage. The light red dashed line indicates the low level of gene connectivity.

are important candidate genes for pig growth and development and average daily weight gain (39). *BTRC* affects spermatogenesis and mammary gland development in mice (40). The four hub genes showed a high degree of interaction. Notably, some non-hub genes, such as *GLI1*, *WIF1*, *DKK2*, *FZD3*, and *WNT4* genes, have been reported to be associated with mammary gland development. This demonstrated that these genes play important roles in animal growth and development.

3.7. Validation of gene expression

For the preliminary validation of the RNA-sequencing data, fifteen DEGs were selected for experimental validation and statistical analysis in another random camel population, including nine upregulated genes (*CAMK2D, CAMK2A, BTRC, PRKAB2, PLCB4, PPP1R3A, GLI1, DKK2,* and *KCNJ12*) and six downregulated genes (*SLC1A1, LTF, WIF1, FZD3, WNT4*, and *PIGR*). The gene expression levels were normalized relative to β-actin. The expression trends of fifteen genes in the random camel population were consistent with the transcriptome results (Figure 7 and Supplementary Table S12).

4. Discussion

Camels, like cattle, have four mammary glands in the inguinal region responsible for the synthesis and secretion of milk proteins (41). Because camels do not have gland cisterns for storage, milk production is a reflex secretion (42). Previous studies have demonstrated that milkability traits can be assessed using udder and teat morphological traits of camels, contributing to the selection of high-producing camels (43). Camel teats have a double ductal anatomy with a promising prophylactic effect against mastitis-causing pathogens (44). Although whole-genome sequencing of dromedaries and Bactrian camels has been completed (45, 46), gene annotation remains incomplete. It is necessary to obtain more data for the further annotation of camelids. To the best of our knowledge, the present study is the first RNA-sequencing study of mammary gland developmental gene expression in young and adult camels. Although the sample sizes in the current study were relatively limited, significant differences were found between the morphology and transcription profiles of mammary gland tissue of young and adult camels. Additionally, the three mammary gland samples in each group showed a high correlation, proving that the biological samples within each group had good repeatability.

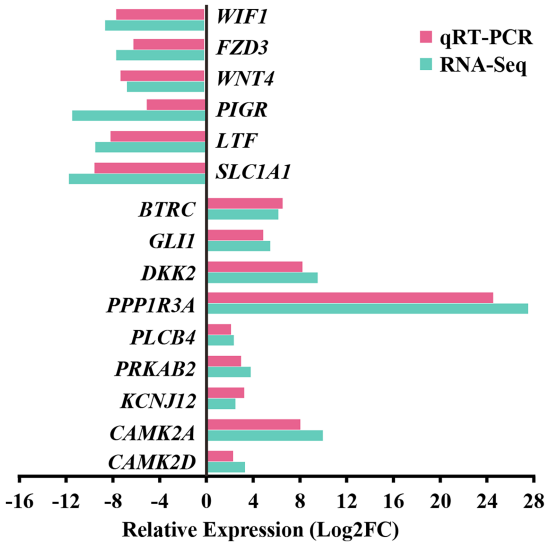


FIGURE 7
Validation of RNA-seq differential gene expression levels using RT-qPCR. The relative expression levels of both the RT-qPCR and RNA-seq were expressed as log2FoldChange (FC), which is defined as the ratio of the expression level in the young camels to that in the adult camels for a specific gene. Error bars are standard deviations across three camels per group evaluated.

The morphology of mammary glands undergoes significant changes during development. A distinctive feature of the mammary gland development is its extensive proliferative and differentiation potential, particularly during puberty and pregnancy. Adolescence is the ideal stage for studying mammary gland development. The mammary gland develops its adult form through a biological process referred to as branching morphogenesis (47). Our observations of the mammary gland are consistent with those of previous studies (44), showing that age and lactation strongly influence the microscopic anatomy of the camel mammary gland.

Extensive experimental efforts have been made to investigate the endocrine signaling and pathways that control the proliferation and differentiation of mammary epithelial cells. Such studies have reported the regulatory roles of multiple signaling pathways in mammary gland proliferation and development. Several studies have revealed that the Wnt pathway is required at the earliest stage of mammary development in the embryo for specification of the mammary placode and initiation of mammary morphogenesis (48) and in postnatal development for proper stem cell maintenance, branching morphogenesis, and mammary alveolar development (49, 50). The Hedgehog signaling pathway is also necessary for mammary gland development and distinguishes the mammary gland from other epidermal appendages (51, 52). Mammary gland development involves complex interactions among various hormones. The mammary gland is composed of a branched ductal system consisting of milk-producing epithelial cells that form ductile tubules surrounded by a layer of myoepithelial cells that contract in response to oxytocin stimulation, thereby releasing milk (53). After the embryonic and prepubertal stages, further development of the mammary glands is highly dependent on hormones such as steroids, which control the development of breast ducts and alveoli (53). The results of RNA sequencing and histomorphological identification of the mammary glands of young and adult camels lend further support to these pathways and hormones regulating mammary gland development in animals.

The functions of genes involved in these pathways require further investigation. Comprehensive PPI network analysis identified ten hub genes. Among these, variations in the *BTRC* gene have been reported to influence spermatogenesis and mammary gland development in mice (40). *PRKAB2* and *PRKAG3* have been implicated in animal growth, slaughter, and meat quality regulation (38). In pigs, *PLCB4* is a candidate gene for growth, development, and meat production (39). Among these non-core genes, *GLI1* regulates the glioma-associated oncogene homolog 1 (*GLI1*) activator and causes mammary bud formation failure in mice (54). The *WIF1* gene is expressed at high levels in normal human breast cells and mammary tissues and maintains mammary gland development (55). *DKK2* and *FZD3* are closely related to the Wnt signaling pathway during embryonic mammary gland development (56). The Wnt ligand *WNT4* plays a critical role in normal mammary gland development. Our study identified a set of crucial genes that might exert their functions by regulating the Wnt, Hedgehog, and sex hormone signaling pathways. However, additional functional experiments are required to confirm this hypothesis.

In summary, this study is of great value for identifying critical genes involved in mammary gland development and can inform subsequent studies on the progression mechanism of mammary gland development in camels. However, there were some limitations. The present study is limited by the small samples size of each group. Three camels were used as replicates at each stage. The findings of the study need to be verified in the future through studies with an expanded sample size. Additionally, there are differences in genetic information between individual animals, which may affect the results of the differential expression analysis of the mammary glands (57). Although distinct differences in the function and structure of mammary glands have been found between young and adult camels, mammary gland growth and development are complex biological processes, and inclusion of more sampling and sequencing time points might help to elucidate the molecular change mechanisms of the underlying biological processes and better reflect the continuous characteristics of dynamic gene changes during mammary gland development.

5. Conclusion

The Bactrian camel is a versatile animal with high ecological and economic value in remote desert areas. Milk synthesis is an essential function of the mammary glands in mammals. This study described the microstructure and transcriptome profiles of the mammary gland parenchyma of young and adult camels. Differential expression analysis revealed 2,851 DEGs, of which 1,420 were upregulated and 1,431 were downregulated. Functional enrichment analysis revealed significant enrichment of the Wnt, Hedgehog, oxytocin, and insulin signaling pathways and steroid biosynthesis. The *PRKAB2*, *PRKAG3*, *PLCB4*, *BTRC*, *GLI1*, *WIF1*, *DKK2*, *FZD3*, and *WNT4* genes involved in these pathways should be considered as important candidate genes for mammary gland development. Nevertheless, the potential functions of these genes still require further verification. This study revealed DEGs and signaling pathways related to mammary gland development in Bactrian camels and laid a theoretical foundation for the improvement of milk production traits in camels.

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: https://www.ncbi.nlm.nih.gov/bioproject; PRJNA946168.

Ethics statement

The animal study was reviewed and approved by Ethics Committee of Xinjiang University. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

JY conceived the study. H Yao writing-original draft and writing-review and editing. H Yao, ZD, and XL conducted conceptualization and writing-review and editing. ZZ, WM, ZH, H Yan, YW, ZH, ZW, and GC collected the experimental samples. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the Key Technology Research and Development Program in Xinjiang Uygur Autonomous Region (2018B01003), the National Key Research and Development Projects of China (2019YFC1606103), and the Postgraduate Scientific Research Innovation Program of Xinjiang Uygur Autonomous Region (XJ2019G026).

Acknowledgments

The authors would like to extend their special thanks to the camel breeders and camel farm managers, as well as to the experts, veterinarians, and laboratories who lent their assistance.

Conflict of interest

GC was employed by Xinjiang Wangyuan Camel Milk Limited Company.

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

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.2023.1196950/full#supplementary-material

References

1. Faye B, Konuspayeva G. *The encounter between bactrian and dromedary camels in central Asia*, vol. *451* Verlag der Österreichischen Akademie der Wissenschaften (2012) 27–33.

2. Yang J, Dou Z, Peng X, Wang H, Shen T, Liu J, et al. Transcriptomics and proteomics analyses of anti-cancer mechanisms of TR35–an active fraction from Xinjiang Bactrian camel milk in esophageal carcinoma cell. *Clin Nutr*. (2019) 38:2349–59. doi: 10.1016/j.clnu.2018.10.013

3. Abdalla E, Anis Ashmawy AEH, Farouk MH, Abd El-Rahman Salama O, Khalil F, Seioudy A. Milk production potential in Maghrebi she-camels. *Small Rumin Res*. (2015) 123:129–35. doi: 10.1016/j.smallrumres.2014.11.004

4. Sun L, Fu Y, Yang Y, Wang X, Cui W, Li D, et al. Genomic analyses reveal evidence of independent evolution, demographic history, and extreme environment adaptation of Tibetan plateau *Agaricus bisporus*. *Front Microbiol*. (2019) 10:1786. doi: 10.3389/fmicb.2019.01786

5. Xuan R, Chao T, Zhao X, Wang A, Chu Y, Li Q, et al. Transcriptome profiling of the nonlactating mammary glands of dairy goats reveals the molecular genetic mechanism of mammary cell remodeling. *J Dairy Sci*. (2022) 105:5238–60. doi: 10.3168/jds.2021-21039

6. Bai X, Zheng Z, Liu B, Ji X, Bai Y, Zhang W. Whole blood transcriptional profiling comparison between different milk yield of Chinese Holstein cows using RNA-seq data. *BMC Genomics*. (2016) 17:512. doi: 10.1186/s12864-016-2901-1

7. Harris HR, Willett WC, Vaidya RL, Michels KB. Adolescent dietary patterns and premenopausal breast cancer incidence. *Carcinogenesis*. (2016) 37:376–84. doi: 10.1093/carcin/bgw023

8. Arora R, Siddaraju NK, Manjunatha SS, Sudarshan S, Fairoze MN, Kumar A, et al. Muscle transcriptome provides the first insight into the dynamics of gene expression with progression of age in sheep. *Sci Rep*. (2021) 11:22360. doi: 10.1038/s41598-021-01848-5

9. Lian Y, Gòdia M, Castelló A, Rodríguez-Gil J, Maestre S, Sanchez A, et al. Characterization of the impact of density gradient centrifugation on the profile of the pig sperm transcriptome by RNA-seq. *Front Vet Sci*. (2021) 8:8. doi: 10.3389/fvets.2021.668158

10. Tucker HLM, Parsons CLM, Ellis S, Rhoads ML, Akers RM. Tamoxifen impairs prepubertal mammary development and alters expression of estrogen receptor α (ESR1) and progesterone receptors (PGR). *Domest Anim Endocrinol*. (2016) 54:95–105. doi: 10.1016/j.domaniend.2015.10.002

11. Velayudhan BT, Huderson BP, McGilliard ML, Jiang H, Ellis SE, Akers RM. Effect of staged ovariectomy on measures of mammary growth and development in prepubertal dairy heifers. *Animal*. (2012) 6:941–51. doi: 10.1017/s1751731111002333

12. Nørgaard JV, Nielsen MO, Theil PK, Sørensen MT, Safayi S, Sejrsen K. Development of mammary glands of fat sheep submitted to restricted feeding during late pregnancy. *Small Rumin Res*. (2008) 76:155–65. doi: 10.1016/j.smallrumres.2007.11.001

13. Hughes K. Comparative mammary gland postnatal development and tumourigenesis in the sheep, cow, cat and rabbit: exploring the menagerie. *Semin Cell Dev Biol*. (2020) 114:186–95. doi: 10.1016/j.semcdb.2020.09.010

14. Williams SA, Harata-Lee Y, Comerford I, Anderson RL, Smyth MJ, McColl SR. Multiple functions of CXCL12 in a syngeneic model of breast cancer. *Mol Cancer*. (2010) 9:250. doi: 10.1186/1476-4598-9-250

15. Hao Z, Zhou H, Hickford JGH, Gong H, Wang J, Hu J, et al. Transcriptome profile analysis of mammary gland tissue from two breeds of lactating sheep. *Genes*. (2019) 10:781. doi: 10.3390/genes10100781

16. Yao H, Liu M, Ma W, Yue H, Su Z, Song R, et al. Prevalence and pathology of Cephalopina titillator infestation in *Camelus bactrianus* from Xinjiang, China. *BMC Vet Res*. (2022) 18:360. doi: 10.1186/s12917-022-03464-5

17. Zhu H, Blum RH, Bernareggi D, Ask EH, Wu Z, Hoel HJ, et al. Metabolic reprograming via deletion of CISH in human iPSC-derived NK cells promotes *in vivo*

persistence and enhances anti-tumor activity. *Cell Stem Cell*. (2020) 27:224–237.e6. doi: 10.1016/j.stem.2020.05.008

18. Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat Commun*. (2021) 12:1165. doi: 10.1038/s41467-021-21450-7

19. Liu G, Yang G, Zhao G, Guo C, Zeng Y, Xue Y, et al. Spatial transcriptomic profiling to identify mesoderm progenitors with precision genomic screening and functional confirmation. *Cell Prolif*. (2022) 55:e13298. doi: 10.1111/cpr.13298

20. Peng H, Guo Q, Xiao Y, Su T, Jiang TJ, Guo LJ, et al. ASPH regulates osteogenic differentiation and cellular senescence of BMSCs. *Front Cell Dev Biol*. (2020) 8:872. doi: 10.3389/fcell.2020.00872

21. Rudolph MC, McManaman JL, Hunter L, Phang T, Neville MC. Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. *J Mammary Gland Biol Neoplasia*. (2003) 8:287–307. doi: 10.1023/b:jomg.0000010030.73983.57

22. Loor JJ, Batistel F, Bionaz M, Hurley WL, Vargas-Bello-Pérez E. Mammary gland: gene networks controlling development and involution* In: PLH McSweeney and JP McNamara, editors. *Encyclopedia of dairy sciences. 3rd* ed. Oxford: Academic Press (2022). 167–74.

23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. (2019) 47:D607–d613. doi: 10.1093/nar/gky1131

24. Nangraj AS, Selvaraj G, Kaliamurthi S, Kaushik A, Cho W, Wei D-Q. Integrated PPI and WGCNA retrieving shared gene signatures between Barrett's esophagus and esophageal adenocarcinoma. *Front Pharmacol*. (2020) 11:11. doi: 10.3389/fphar.2020.00881

25. Yu X, Wu Y, Zhang J, Jirimutu ZA, Chen J. Pre-evaluation of humoral immune response of Bactrian camels by the quantification of Th2 cytokines using real-time PCR. *J Biomed Res*. (2020) 34:387–94. doi: 10.7555/jbr.34.20190035

26. Liu Z, An L, Lin S, Wu T, Li X, Tu J, et al. Comparative physiological and transcriptomic analysis of pear leaves under distinct training systems. *Sci Rep*. (2020) 10:18892. doi: 10.1038/s41598-020-75794-z

27. Hindman AR, Mo XM, Helber HL, Kovalchin CE, Ravichandran N, Murphy AR, et al. Varying susceptibility of the female mammary gland to in utero windows of BPA exposure. *Endocrinology*. (2017) 158:3435–47. doi: 10.1210/en.2017-00116

28. Finot L, Chanat E, Dessauge F. Molecular signature of the putative stem/progenitor cells committed to the development of the bovine mammary gland at puberty. *Sci Rep*. (2018) 8:16194. doi: 10.1038/s41598-018-34691-2

29. Chatterjee SJ, Halaoui R, Deagle RC, Rejon C, McCaffrey L. Numb regulates cell tension required for mammary duct elongation. *Biol Open*. (2019) 8:bio042341. doi: 10.1242/bio.042341

30. Gerashchenko TS, Zolotaryova SY, Kiselev AM, Tashireva LA, Novikov NM, Krakhmal NV, et al. The activity of KIF14, Mieap, and EZR in a new type of the invasive component, Torpedo-like structures, predetermines the metastatic potential of breast Cancer. *Cancers*. (2020) 12:1909. doi: 10.3390/cancers12071909

31. Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cells*. (2014) 159:1110–25. doi: 10.1016/j.cell.2014.10.013

32. Shahzad N, Munir T, Javed M, Tasneem F, Aslam B, Ali M, et al. SHISA3, an antagonist of the Wnt/β-catenin signaling, is epigenetically silenced and its ectopic expression suppresses growth in breast cancer. *PLoS One*. (2020) 15:e0236192. doi: 10.1371/journal.pone.0236192

33. Jia S, Zhou J, Wee Y, Mikkola ML, Schneider P, D'Souza RN. Anti-EDAR agonist antibody therapy resolves palate defects in *Pax9*$^{-/-}$ mice. *J Dent Res*. (2017) 96:1282–9. doi: 10.1177/0022034517726073

34. Sun G, Wang C, Wang S, Sun H, Zeng K, Zou R, et al. An H3K4me3 reader, BAP18 as an adaptor of COMPASS-like core subunits co-activates ERα action and associates with the sensitivity of antiestrogen in breast cancer. *Nucleic Acids Res*. (2020) 48:10768–84. doi: 10.1093/nar/gkaa787

35. Reynolds P, Canchola AJ, Duffy CN, Hurley S, Neuhausen SL, Horn-Ross PL, et al. Urinary cadmium and timing of menarche and pubertal development in girls. *Environ Res*. (2020) 183:109224. doi: 10.1016/j.envres.2020.109224

36. Zhang Y, Gc S, Patel SB, Liu Y, Paterson AJ, Kappes JC, et al. Growth hormone (GH) receptor (GHR)-specific inhibition of GH-induced signaling by soluble IGF-1 receptor (sol IGF-1R). *Mol Cell Endocrinol*. (2019) 492:110445. doi: 10.1016/j.mce.2019.05.004

37. Dianati E, Poiraud J, Weber-Ouellette A, Plante I. Connexins, E-cadherin, claudin-7 and β-catenin transiently form junctional nexuses during the post-natal mammary gland development. *Dev Biol*. (2016) 416:52–68. doi: 10.1016/j.ydbio.2016.06.011

38. Milan D, Jeon J-T, Looft C, Amarger V, Robic A, Thelander M, et al. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science*. (2000) 288:1248–51. doi: 10.1126/science.288.5469.1248

39. Edea Z, Hong JK, Jung JH, Kim DW, Kim YM, Kim ES, et al. Detecting selection signatures between Duroc and Duroc synthetic pig populations using high-density SNP chip. *Anim Genet*. (2017) 48:473–7. doi: 10.1111/age.12559

40. Cortellari M, Barbato M, Talenti A, Bionda A, Carta A, Ciampolini R, et al. The climatic and genetic heritage of Italian goat breeds with genomic SNP data. *Sci Rep*. (2021) 11:10986. doi: 10.1038/s41598-021-89900-2

41. Yang Y, Fang X, Yang R, Yu H, Jiang P, Sun B, et al. MiR-152 regulates apoptosis and triglyceride production in MECs via targeting ACAA2 and HSD17B12 genes. *Sci Rep*. (2018) 8:417. doi: 10.1038/s41598-017-18804-x

42. Alluwaimi A. The Camel's (*Camelus dromedarius*) mammary gland immune system in health and disease. *Adv Dairy Res*. (2017) 05:5. doi: 10.4172/2329-888X.1000171

43. Kumar M, Nehara M, Prakash V, Pannu U, Jyotsana B. Udder, teat, and milk vein measurements of Indian dromedary camel and its relationship with milkability traits. *Trop Anim Health Prod*. (2023) 55:36. doi: 10.1007/s11250-023-03457-y

44. Kausar R. Gross and microscopic anatomy of mammary gland of dromedaries under different physiological conditions. *Pak Vet J*. (2001) 21:189–93.

45. Jirimutu WZ, Ding G, Chen G, Sun Y, Sun Z, Zhang H, et al. The Bactrian camels genome sequencing and analysis consortium. correction: corrigendum: genome sequences of wild and domestic bactrian camels. *Nat Commun*. (2013) 4:2089. doi: 10.1038/ncomms3089

46. Fitak RR, Mohandesan E, Corander J, Burger PA. The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol Ecol Resour*. (2016) 16:314–24. doi: 10.1111/1755-0998.12443

47. Sieh W, Rothstein JH, Klein RJ, Alexeeff SE, Sakoda LC, Jorgenson E, et al. Identification of 31 loci for mammographic density phenotypes and their associations with breast cancer risk. *Nat Commun*. (2020) 11:5116. doi: 10.1038/s41467-020-18883-x

48. Roarty K, Shore AN, Creighton CJ, Rosen JM. Ror2 regulates branching, differentiation, and actin-cytoskeletal dynamics within the mammary epithelium. *J Cell Biol*. (2015) 208:351–66. doi: 10.1083/jcb.201408058

49. Watson CJ, Khaled WT. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development*. (2008) 135:995–1003. doi: 10.1242/dev.005439

50. Szmatoła T, Gurgul A, Jasielczuk I, Ząbek T, Ropka-Molik K, Litwińczuk Z, et al. A comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. *Animals*. (2019) 9:1024. doi: 10.3390/ani9121024

51. Hatsell SJ, Cowin P. Gli3-mediated repression of hedgehog targets is required for normal mammary development. *Development*. (2006) 133:3661–70. doi: 10.1242/dev.02542

52. Behbod F, Rosen JM. Will cancer stem cells provide new therapeutic targets? *Carcinogenesis*. (2005) 26:703–11. doi: 10.1093/carcin/bgh293

53. Hennighausen L, Robinson GW. Signaling pathways in mammary gland development. *Dev Cell*. (2001) 1:467–75. doi: 10.1016/S1534-5807(01)00064-8

54. Niyaz M, Khan MS, Mudassar S. Hedgehog signaling: An Achilles' heel in Cancer. *Transl Oncol*. (2019) 12:1334–44. doi: 10.1016/j.tranon.2019.07.004

55. Wang N, Wang Z, Wang Y, Xie X, Shen J, Peng C, et al. Dietary compound isoliquiritigenin prevents mammary carcinogenesis by inhibiting breast cancer stem cells through WIF1 demethylation. *Oncotarget*. (2015) 6:9854–76. doi: 10.18632/oncotarget.3396

56. Pausch H, Jung S, Edel C, Emmerling R, Krogmeier D, Götz KU, et al. Genome-wide association study uncovers four QTL predisposing to supernumerary teats in cattle. *Anim Genet*. (2012) 43:689–95. doi: 10.1111/j.1365-2052.2012.02340.x

57. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. (2016) 17:13. doi: 10.1186/s13059-016-0881-8

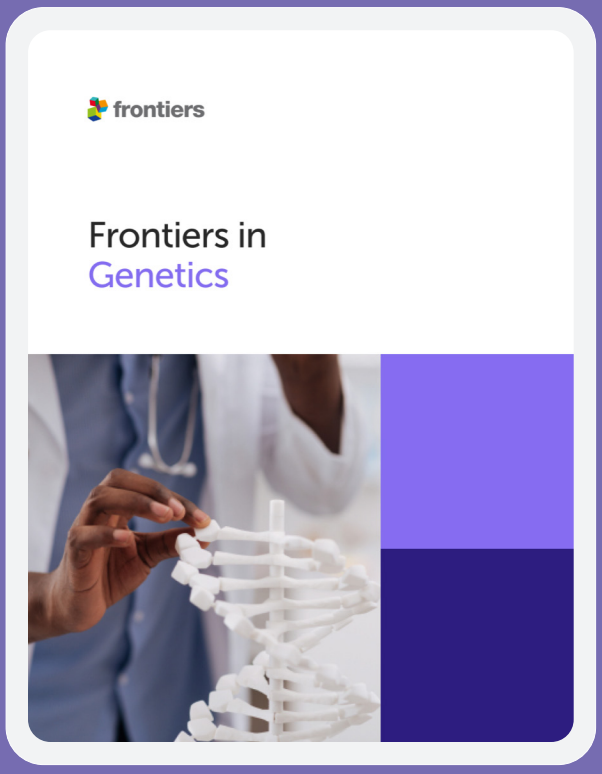
Frontiers in Genetics

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