

Genomic insights into sheep and goat breeding efficiency

Edited by
Zhihong Liu and Fei Hao

Published in
Frontiers in Veterinary Science
Frontiers in Genetics



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 978-2-8325-7231-3
DOI 10.3389/978-2-8325-7231-3

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Genomic insights into sheep and goat breeding efficiency

Topic editors

Zhihong Liu — Inner Mongolia Agricultural University, China

Fei Hao — Northumbria University, United Kingdom

Citation

Liu, Z., Hao, F., eds. (2025). *Genomic insights into sheep and goat breeding efficiency*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7231-3

Table of contents

- 05 **Editorial: Genomic insights into sheep and goat breeding efficiency**
Fei Hao, Mingli Peng, Ziyang Xu, Xiao Zhang, Jiawei Wang and Rui Ding
- 08 **Genomic prediction based on a joint reference population for the Xinjiang Brown cattle**
Menghua Zhang, Lei Xu, Haibo Lu, Hanpeng Luo, Jinghang Zhou, Dan Wang, Xiaoxue Zhang, Xixia Huang and Yachun Wang
- 26 **Effects of nonsynonymous single nucleotide polymorphisms of the *KIAA1217*, *SNTA1* and *LTBP1* genes on the growth traits of Ujumqin sheep**
Zhichen Liu, Qing Qin, Chongyan Zhang, Xiaolong Xu, Dongliang Dai, Mingxi Lan, Yichuan Wang, Jingwen Zhang, Dan Zhao, Deqing Kong, Tian Qin, Danni Wu, Xuedan Gong, Xingyu Zhou, Alatan Suhe, Zhixin Wang and Zhihong Liu
- 36 **Whole-genome resequencing reveals genetic diversity, differentiation, and selection signatures of yak breeds/populations in southwestern China**
Shilin Zhang, Jing Li, Yanhua Zhao, Yujun Tang, Hao Li, Tianzeng Song, Tianwu An, Jiuqiang Guan, Xiaowei Li and Ming Zhang
- 46 **Transcriptome analysis revealed the characteristics and functions of long non-coding RNAs in the hypothalamus during sexual maturation in goats**
Qing Li, Tianle Chao, Yanyan Wang, Rong Xuan, Yanfei Guo, Peipei He, Lu Zhang and Jianmin Wang
- 56 **Identification and functional prediction of lncRNAs associated with intramuscular lipid deposition in Guangling donkeys**
Yongdong Peng, Mingzhu Zhu, Yunyi Gong and Changfa Wang
- 66 **Combining multiomics to analyze the molecular mechanism of hair follicle cycle change in cashmere goats from Inner Mongolia**
Chongyan Zhang, Qing Qin, Zhichen Liu, Yichuan Wang, Mingxi Lan, Dan Zhao, Jingwen Zhang, Zhixin Wang, Jinquan Li and Zhihong Liu
- 77 **Analysis of genomic copy number variations through whole-genome scan in Yunling cattle**
Dong Dang, Lilian Zhang, Lutao Gao, Lin Peng, Jian Chen and Linnan Yang
- 85 **Selection signature analysis reveals genes associated with tail phenotype in sheep**
Yunxia Qi, Xiaolong He, Biao Wang, Chaoyun Yang, Lai Da, Bin Liu, Wenguang Zhang, Shaoyin Fu and Yongbin Liu

- 96 **Unlocking the genetic secrets of Dorper sheep: insights into wool shedding and hair follicle development**
Xiaochun Yuan, Ke Meng, Yayan Wang, Yifan Wang, Cuili Pan, Haoran Sun, Jankui Wang and Xinhai Li
- 112 **Identification of selective sweep and associated QTL traits in Iranian *Ovis aries* and *Ovis orientalis* populations**
Sadegh Taheri, Ali Javadmanesh and Saeed Zerehdaran
- 124 **Comparative proteomic analysis of the telogen-to-anagen transition in cashmere goat secondary hair follicles**
Xiaoyu Han, Guizhen Gao, Nanxiang Sun, Bai Dai, Liqing Ren, Haobing Bai, Jianing Liu, Jie Liu, Hongyan Zhao and Dongjun Liu
- 138 **Genetic diversity, reproductive performance, and genetic enhancement strategies in Huang-Huai goats**
Kai Quan, Huibin Shi, Caihong Wei, Jun Li, Kun Liu, Huihua Wang, Wei Sun and Haoyuan Han
- 147 **The genetic diversity and population structure of native horse breeds in Xinjiang, China**
Chi Tang, Baoyu Yang, Gulibaheti Dawulietihan, Li Xue, Shuyuan Liu, Yinamujiang Yalimaimaiti, Qingzheng Wang, Na Yang, Xiaoyuan Sun, Yaru Wang, Ailifeire Wumaier, Serik Khizat, Tolegen Assanbayev, Zhassulan Kozhanov, Kursantbek Attokurov, Elmurat Obdunov, Hangsen Li, Aikebaier Rehemani, Xiaoling Zhou, Wumaierjiang Aizimu, Kairat Iskhan and Gemingguli Muhatai



OPEN ACCESS

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

*CORRESPONDENCE

Fei Hao
✉ feihao@imu.edu.cn

RECEIVED 01 October 2025

ACCEPTED 21 October 2025

PUBLISHED 04 November 2025

CITATION

Hao F, Peng M, Xu Z, Zhang X, Wang J and
Ding R (2025) Editorial: Genomic insights into
sheep and goat breeding efficiency.
Front. Vet. Sci. 12:1717356.
doi: 10.3389/fvets.2025.1717356

COPYRIGHT

© 2025 Hao, Peng, Xu, Zhang, Wang and
Ding. 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: Genomic insights into sheep and goat breeding efficiency

Fei Hao^{1,2*}, Mingli Peng¹, Ziyang Xu¹, Xiao Zhang¹, Jiawei Wang¹
and Rui Ding¹

¹Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot, China, ²Northumbria University, Newcastle upon Tyne, United Kingdom,

KEYWORDS

Grassland Livestock, genomic, gene, phenotype, breeding efficiency

Editorial on the Research Topic

Genomic insights into sheep and goat breeding efficiency

Grassland Livestock, as the gene, phenotype, Breeding efficiency and ecosystem, provide high-quality livestock products, such as meat, milk and down, that are essential for human survival, maintain the energy flow between forage and humans, and also carry the historical connection between thousands of years of nomadic civilization and the transformation of modern animal husbandry (1–4). With the development of cell biology, molecular biology, and genomics, researchers have conducted extensive studies on the periodic growth of cashmere goat hair follicles, reproductive efficiency of sheep, and improvements in dairy cow production (5, 6). This scaffolds a resolution to the dichotomy between “protecting grassland ecology” and “improving livestock productivity.” Such related research is a core focus in animal husbandry, veterinary science, genetics, and ecology (7).

In recent years, research on grazing livestock has progressed from traditional phenotypic observations and population surveys to a new stage of elucidating the molecular mechanisms of phenotypes and molecularly-informed breeding (8). In-depth research has been conducted on key economic traits, such as growth, reproduction, disease prevention, and the control of grazing livestock (9). Related research provides scientific solutions to the problems faced in grassland animal husbandry, such as long breeding cycles, difficult trait improvements, and insufficient ecological adaptability (10).

Consequently, *Frontiers in Veterinary Science* has established a Research Topic, “Genomic Insights into Sheep and Goat Breeding Efficiency,” carefully selecting 13 high-quality, original research articles focusing on the field of grazing livestock. These involve important grazing animals such as sheep, goats, yaks, Yunling cattle, and donkeys, and cover the complete research chain from “genetic resource evaluation to functional gene mining to production trait regulation to molecular breeding application.” These papers cover the molecular basis of important biological characteristics of grazing livestock from different perspectives, providing references for promoting innovation in grazing livestock scientific research and promoting sustainable commercial development.

In terms of genetic resources and adaptive evolutionary research in grazing livestock, multiple studies have focused on the genomic characteristics of typical livestock species, providing a molecular basis for the conservation of genetic diversity and stress-resistant

breeding. Tang et al. used 10X whole-genome sequencing to conduct single nucleotide polymorphism analysis on seven local horse breeds in Xinjiang, China, and found high population genetic diversity among these local breeds. Marked genetic differences from other horse breeds from Europe, central Asia, western Asia, and China were reported, elucidating differences in distribution patterns, evolutionary characteristics, and genetic diversity.

Zhang S. et al. conducted whole-genome resequencing of six yak populations in south-western China and found rich genetic diversity in yaks from this region. Tibetan yaks showed lower nucleotide diversity because of geographical isolation, whereas Muli yaks were substantially different from the other groups. Strong candidate genes related to high-altitude adaptation, growth, and development were found, addressing a gap in genomic research on yak populations in south-western China.

Dang et al. successfully constructed a genomic copy number variation (CNV) map for Yunling cattle that will facilitate an in-depth analysis of the genetic mechanisms underlying the formation of economic traits, such as subcutaneous fat thickness and *longissimus dorsi* muscle area in Yunling cattle.

By comparing selective sweep signals between Iranian domestic sheep and wild Mouflon sheep, Taheri et al. found that genes, such as *ADGRB3* and *CAPN2*, in domestic sheep were strongly correlated with economic traits, such as body weight and milk yield, whereas genes such as *ACAN* and *MGST3* in wild sheep were related to adaptive traits, such as daily weight gain and bone weight. This indicates the differential effects of artificial and natural selection on the sheep genome, providing a new perspective for the utilization of sheep genetic resources.

Another core focus of this Research Topic is the regulatory molecular mechanisms of important production traits. Multiple papers in this Research Topic target key traits, such as hair follicle development, reproductive performance, growth, and meat quality, mining many functional genes and regulatory pathways with potential application in the field. In hair follicle development research, Han et al. used proteomics techniques to analyze differences between the telogen and anagen phases of secondary hair follicles in cashmere goats. They found that *ADAM17*, *SFRP1*, and *PPP1CA* proteins might promote hair follicle cycle transition by regulating signaling pathways, such as Notch and Wnt. Zhang C. et al. used multi-omics joint analysis techniques to elucidate the important role of ribosomal proteins during the hair follicle cycle transition. Yuan et al., focusing on the depilation characteristics of Dorper sheep, screened hair follicle development-related genes, such as *DBI*, *FZD3*, and *ZDHHC21*, providing new targets for the regulation of the sheep hair follicle cycle. Related research provides references for studying the mechanisms of hair follicle growth and development in goats and sheep.

In the field of reproductive efficiency traits, Quan et al. systematically analyzed the reproductive performance of Huanghuai goats and found an average litter size of 2.74 and an annual reproductive rate of 418.96%. Combined with transcriptome sequencing technology, they screened candidate genes, such as *PTX3* and *MMP13*, for prolificacy traits.

In the field of meat yield and meat quality-related traits, Liu et al. genotyped non-synonymous single nucleotide polymorphisms of three candidate genes (*KIAA1217*, *SNTA1*, and *LTBP1*) in Ujimqin sheep and performed association analysis

between the genotyping results and growth traits. These results provide important data for genetic marker-assisted selection of Ujimqin sheep.

Peng et al., using transcriptome data from the *longissimus dorsi* muscle of Guangling donkeys, analyzed the importance of long non-coding RNAs in intramuscular fat deposition, expanding the research dimension of non-coding RNAs in the regulation of grazing livestock traits.

Research in molecular breeding technology innovation and application has focused on addressing the bottlenecks in breeding in small populations and optimizing trait prediction models to provide new methods for improving the breeding efficiency of grazing livestock. Zhang S. et al. analyzed the genetic structure of 485 Xinjiang Brown cattle and 2,633 Chinese Holstein cattle to establish a cross-bred joint reference population. They also evaluated the estimated genomic breeding values of milk production traits in Xinjiang Brown cattle. This study provides basic data for genomic prediction and selection in dairy cattle. Qi et al. screened candidate genes related to tail-type using whole-genome re-sequencing in Mongolian (short, fat-tailed) and Bamei (long, thin-tailed) mutton sheep. Of these, *PDGFD*, *GLIS1*, and *VRTN* were strongly associated with tail-fat deposition and tail length. This study provides a new theoretical basis for the molecular breeding of tail-type traits in sheep.

Li et al. conducted transcriptome sequencing of the hypothalamus of Jining Gray goats, identifying 237 differentially-expressed, long non-coding RNAs and analyzed their regulatory role in goat sexual maturity, providing a reference for improving goat reproductive efficiency.

In summary, the 13 research reports in this Research Topic comprehensively present the current research frontiers in the field of grazing livestock, including an in-depth analysis of the genetic resources of livestock species, such as yaks and Yunling cattle; innovative exploration of the regulatory mechanisms of key traits, such as hair follicles and reproduction; and practical breakthroughs in technologies, such as cross-breeding. These achievements not only enrich the theoretical system of grazing livestock biology but also provide scientific support for the synergistic advancement of animal health, ecological protection, and commercial development under the “One Health” concept.

Author contributions

FH: Writing – original draft, Writing – review & editing. MP: Writing – original draft, Writing – review & editing. ZX: Writing – original draft, Writing – review & editing. XZ: Writing – original draft, Writing – review & editing. JW: Writing – original draft, Writing – review & editing. RD: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was financially supported by National Natural Science Foundation of China Project Study on the Regulatory Mechanism of the EDAR

Gene in the Growth and Development of Secondary Hair Follicles in Cashmere Goats (Project No. 32360811), State Key Laboratory of Reproductive Regulation & Breeding of Grassland Livestock (Grant No. 2025KYPT0081).

Acknowledgments

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

Conflict of interest

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

References

1. Soca P, Do Carmo M, Paparamborda I, Figueroa V, Scarlato S, Ruggia A, et al. Review: a hierarchical research model to foster dialog between grazing ecology and beef cow energetics to support ecological intensification of native grassland. *Animal*. (2024) 17:101372. doi: 10.1016/j.animal.2024.101372
2. Henchion M, Moloney AP, Hyland J, Zimmermann J, McCarthy S. Review: Trends for meat, milk and egg consumption for the next decades and the role played by livestock systems in the global production of proteins. *Animal*. (2021) 15:100287. doi: 10.1016/j.animal.2021.100287
3. Lupton CJ. Prospects for expanded mohair and cashmere production and processing in the United States of America. *J Anim Sci*. (1996) 74:1164–72. doi: 10.2527/1996.7451164x
4. Maestre FT, Le Bagousse-Pinguet Y, Delgado-Baquerizo M, Eldridge DJ, Saiz H, Berdugo M, et al. Grazing and ecosystem service delivery in global drylands. *Science*. (2022) 378:915–20. doi: 10.1126/science.abq4062
5. Chunhua Z, Le F, Shengli L, Sachula W, Bao H, Lan M, et al. Multi-omics and AI-driven advances in miRNA-mediated hair follicle regulation in cashmere goats. *Front Vet Sci*. (2025) 12:1635202. doi: 10.3389/fvets.2025.1635202
6. Jiang X, Wang L. Grassland-based ruminant farming systems in China: potential, challenges and a way forward. *Anim Nutr*. (2022) 10:243–8. doi: 10.1016/j.aninu.2022.04.007
7. Xu S, Akhatayeva Z, Liu J, Feng X, Yu Y, Badaoui B, et al. Genetic advancements and future directions in ruminant livestock breeding: from reference genomes to multiomics innovations. *Sci China Life Sci*. (2025) 68:934–60. doi: 10.1007/s11427-024-2744-4
8. Wang HT, Li TT, Huang X, Ma RL, Liu QY. Application of genetic modification technologies in molecular design breeding of sheep. *Yi Chuan*. (2021) 43:580–600. doi: 10.11813/j.hered.2021.4306-580
9. Gunasekera U, VanderWaal K, Arzt J, Perez A. Foot-and-mouth disease reproduction number: a scoping review. *Front Vet Sci*. (2025) 12:1576974. doi: 10.3389/fvets.2025.1576974
10. Radchuk V, Reed T, Teplitsky C, van de Pol M, Charmantier A, Hassall C, et al. Adaptive responses of animals to climate change are most likely insufficient. *Nat Commun*. (2019) 10:3109. doi: 10.1038/s41467-019-10924-4

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Mao Yongjiang,
Yangzhou University, China
Linyang Xu,
Chinese Academy of Agricultural Sciences,
China

*CORRESPONDENCE

Xixia Huang,
✉ au-huangxixia@163.com
Yachun Wang,
✉ wangyachun@cau.edu.cn

RECEIVED 01 March 2024

ACCEPTED 10 April 2024

PUBLISHED 26 April 2024

CITATION

Zhang M, Xu L, Lu H, Luo H, Zhou J, Wang D,
Zhang X, Huang X and Wang Y (2024), Genomic
prediction based on a joint reference
population for the Xinjiang Brown cattle.
Front. Genet. 15:1394636.
doi: 10.3389/fgene.2024.1394636

COPYRIGHT

© 2024 Zhang, Xu, Lu, Luo, Zhou, Wang, Zhang,
Huang 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.

Genomic prediction based on a joint reference population for the Xinjiang Brown cattle

Menghua Zhang¹, Lei Xu¹, Haibo Lu², Hanpeng Luo²,
Jinghang Zhou³, Dan Wang¹, Xiaoxue Zhang¹, Xixia Huang^{1*} and
Yachun Wang^{2*}

¹College of Animal Science, Xinjiang Agricultural University, Urumqi, China, ²Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture of China, National Engineering Laboratory of Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, ³Shijiazhuang Molbreeding Biotechnology Co., Ltd., Shijiazhuang, China

Introduction: Xinjiang Brown cattle constitute the largest breed of cattle in Xinjiang. Therefore, it is crucial to establish a genomic evaluation system, especially for those with low levels of breed improvement.

Methods: This study aimed to establish a cross breed joint reference population by analyzing the genetic structure of 485 Xinjiang Brown cattle and 2,633 Chinese Holstein cattle (Illumina GeneSeek GGP bovine 150 K chip). The Bayes method single-step genome-wide best linear unbiased prediction was used to conduct a genomic evaluation of the joint reference population for the milk traits of Xinjiang Brown cattle. The reference population of Chinese Holstein cattle was randomly divided into groups to construct the joint reference population. By comparing the prediction accuracy, estimation bias, and inflation coefficient of the validation population, the optimal number of joint reference populations was determined.

Results and Discussion: The results indicated a distinct genetic structure difference between the two breeds of adult cows, and both breeds should be considered when constructing multi-breed joint reference and validation populations. The reliability range of genome prediction of milk traits in the joint reference population was 0.142–0.465. Initially, it was determined that the inclusion of 600 and 900 Chinese Holstein cattle in the joint reference population positively impacted the genomic prediction of Xinjiang Brown cattle to certain extent. It was feasible to incorporate the Chinese Holstein into Xinjiang Brown cattle population to form a joint reference population for multi-breed genomic evaluation. However, for different Xinjiang Brown cattle populations, a fixed number of Chinese Holstein cattle cannot be directly added during multi-breed genomic selection. Pre-evaluation analysis based on the genetic structure, kinship, and other factors of the current population is required to ensure the authenticity and reliability of genomic predictions and improve estimation accuracy.

KEYWORDS

Xinjiang Brown cattle, multi-breed, genomic prediction, bayes, single-step GBLUP

1 Introduction

Xinjiang Brown cattle is a major breed supporting the development of the cattle industry in Xinjiang, it was the first breed of cattle used for milk and meat purposes after the founding of the People's Republic of China. In 2023, the number of Xinjiang Brown cattle in stock reached 1.16 million; however, the level of breed improvement was low, with a

performance measurement population of <10,000. Therefore, it is important to establish an efficient genomic evaluation system for Xinjiang Brown cattle to improve their genetic level. The application of genome selection technology has significantly enhanced the efficiency of genomic evaluation (Hayes et al., 2016). Because of the implementation of the genome selection for Chinese Holstein cattle in 2008, early and accurate selection of calves and young cattle has been achieved (George et al., 2017), leading to higher accuracy in genomic evaluation and more precise assessment of individual breeding value (Weigel et al., 2010; Dasonneville et al., 2011). In addition, due to early selection and higher accuracy, the rate of genetic progress has doubled (Weller et al., 2017), improving breeding profitability and significantly reducing breeding costs. Although genome selection has been successfully applied to Chinese Holstein cattle population, the low level of production performance measurement and small population size of Xinjiang Brown cattle have hindered the application of genome selection technology. To improve the reliability of genomic predictions, especially for smaller populations, many feasible methods have been proposed, including increasing marker density, constructing linkage disequilibrium (LD) with more markers and causal mutations, and simulation data analysis (de Roos et al., 2009). Simulation and real data analyses (Brøndum et al., 2015) have shown that genomic prediction can play an important role in different populations.

For genome selection, it is necessary to have a reference population with sufficient size and an appropriate genetic structure that simultaneously incorporates genomic and phenotypic information to accurately predict genome estimated breeding values (GEBVs) (Metta et al., 2004; Boichard et al., 2016). Genome selection has recently been widely used in dairy cattle breeding programs. However, its application is limited to populations with a small number of breeds. Establishing a sufficiently large reference population is the most limiting factor for the accurate estimation of SNP effects (Boichard et al., 2016). When conducting genome selection for small populations, the most direct approach to enhancing its reliability is to expand the reference population. Many countries have found effective solutions through international cooperation, leading to joint genomic evaluations (Lund et al., 2011). By connecting France, Germany, Austria, Italy, Slovenia, Switzerland, and the United States of America to the InterGenomics consortium operated by the Interbull Center (Zumbach et al., 2010; Jorjani et al., 2012), genome-wide joint evaluations have been conducted for Brown Swiss bulls and Simmental cattle in Germany and Austria (Edel et al., 2011). Research has shown that by combining different populations of the same breed or related breeds in the reference population, more effective information can be obtained for estimating marker effects. Therefore, more accurate breeding predictions can be obtained from genomic predictions. Accuracy is improved when three related dairy cattle populations, Danish Red, Swedish Red, and Finnish Ayrshire, are combined into a single reference population (Zhou et al., 2014). When four European Holstein populations were combined into a reference population, the reliability increased by 10% (Lund et al., 2011). By combining six Brown Swiss populations, the reliability increased from 6% to 45% (Jorjani et al., 2012). However, multi-breed genomic

evaluation of Xinjiang Brown cattle has not yet been conducted, limiting the optimized utilization of genomic selection technology in their genomic evaluation.

Based on the research foundation for domestic and international multi-breed joint genomic evaluation (Pryce et al., 2011; Lund et al., 2014; Steyn et al., 2019; Xu et al., 2019; Palombo et al., 2021), we proposed to integrate Xinjiang Brown cattle and Chinese Holstein cattle to construct a joint reference population for genome selection. In order to expand the Xinjiang brown cattle genome selection reference group, so as to apply multi-breed genome selection in Xinjiang brown cattle population to improve the prediction reliability. This study aimed to analyze the genetic structures of Xinjiang Brown and Chinese Holstein cattle to establish a multi-breed joint reference population. Using a dual-trait single-step genome-wide best linear unbiased prediction (ssGBLUP) approach, we established a genomic evaluation system for the primary lactation traits of Xinjiang Brown cattle, leveraging the joint reference population of Xinjiang Brown and Chinese Holstein cattle. This improves the accuracy of genomic selection for Xinjiang Brown cattle, creating a core breeding herd of genetically superior dairy Xinjiang Brown cows. Consequently, the genetic improvement of Xinjiang Brown cattle population will be expedited, leading to enhanced genetic levels across the breed.

2 Materials and methods

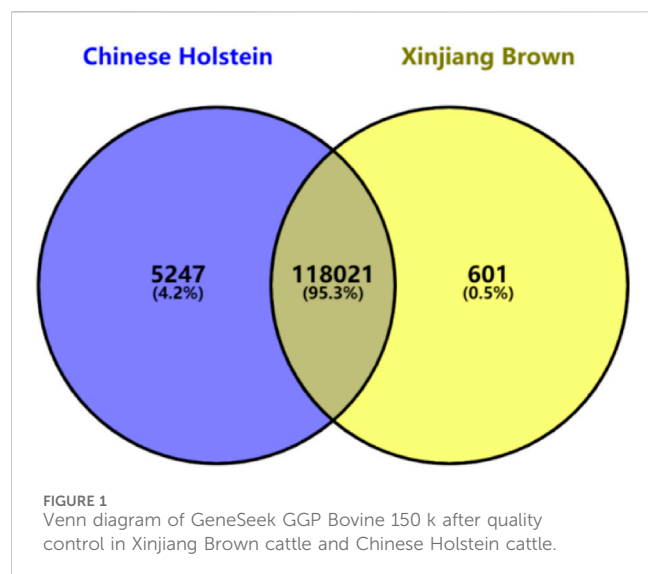
2.1 Sample collection and DNA extraction

A total of 1,729 blood samples were collected from the tail vein of Xinjiang Brown cattle and added to 10 mL EDTA anticoagulant tubes. The samples were then aliquoted into 1.5 mL centrifuge tubes and stored at -20°C . In addition, 66 frozen semen samples were collected from Xinjiang Brown and Brown Swiss bulls used for the artificial insemination of Xinjiang Brown cattle after 1983.

DNA was extracted from the above samples, and the concentration and purity of the obtained genomic DNA were measured using a NanoDrop 2000c spectrophotometer. The OD260/OD280 ratio was 1.7–1.9, indicating good DNA quality. After assessing DNA concentration, purity, and integrity, the samples were stored at -20°C (Ma, 2015).

2.2 Sample screening and chip detection

Phenotypically complete Xinjiang Brown cattle were screened from various Xinjiang Brown cattle farms for chip detection. After screening, 403 cows and 82 bulls from four core farms in Xinjiang region were selected. Moreover, we included 174 Xinjiang Brown cows from Xinjiang Uygur Autonomous Region State-owned Urumqi breeding farm, 50 Xinjiang Brown cows from Xinjiang Tianshan Animal Husbandry Bioengineering Co., Ltd. breeding farm, 130 Xinjiang Brown cows from the Tacheng Agriculture and Animal Husbandry Technology Co., Ltd., 49 Xinjiang Brown cows from Yili New Brown breeding farm, 71 bulls and 11 Brown Swiss bulls from Xinjiang Tianshan Animal Husbandry Bioengineering Co., Ltd. Bull breeding station. Chip data for Chinese Holstein cows were obtained from 2,633 animals in



Beijing, distributed across 18 farms in the region. All of these animals were detected using the Illumina GeneSeek GGP bovine 150 K chip.

2.3 Chip imputation and quality control

A total of 139,376 and 138,892 SNP markers were detected using Xinjiang Brown and Chinese Holstein cattle chip assays, respectively. These data were imputed using Beagle 4.1 software, which infers haplotypes present in the population based on the principle of linkage disequilibrium. To ensure the accuracy of imputation, quality control measures were applied to the chip data.

The quality control criteria were as follows: individuals with a genotyping call rate of <90% were excluded. Only SNPs on chromosomes 1–30 were retained, with an individual genotype missing rate of <10%. SNPs with a minor allele frequency of >0.01 and a Hardy–Weinberg equilibrium p -value $>1 \times 10^{-6}$ were also included. After quality control using the PLINK software, the SNP genotypes were converted to a 0, 1, and 2 format. Finally, 118,622 and 123,268 SNP markers on the autosomal chromosomes of Xinjiang Brown and Chinese Holstein cattle were retained, respectively. Because the number of SNP markers differed between the two breeds after quality control, an intersection of the SNP markers was taken, which resulted in 118,021 common SNP markers for both breeds (Figure 1).

2.4 Genetic structure analysis

2.4.1 Linkage disequilibrium analysis

The .map and .ped files for both breeds were converted to .vcf format using PLINK. The PopLDdecay software was then used to analyze and plot LD decay graphs (<https://github.com/BGI-shenzhen/PopLDdecay>) (Zhang et al., 2019). The LD metric, r^2 , was calculated for the four populations (Hill, 1974). The mean r^2 value was computed at various marker distances of 1 Kb to demonstrate the degree of LD decay across different populations.

2.4.2 Population structure analysis

To infer ancestral populations based on the allele frequencies of descendant individuals, an unsupervised algorithm was employed (Consortium et al., 2009). In this study, genome-wide SNP data were used to calculate the population structure for ancestral admixture components with K values of 2–4 using admixture (Alexander et al., 2009). Visualization of the population structure was performed using the R package “pophelper”.

Analysis was conducted using the FastTree software (<http://www.microbesonline.org/fasttree/>), with the maximum likelihood method adopted for estimation. The Jukes–Cantor + CAT model was used as the default model for nucleotide phylogeny. The credibility of the phylogenetic tree branches was tested using 1,000 bootstrap replicates. Finally, the FigTree software was used for visualization.

2.4.3 Principal component analysis

The Gmatrix package in R was used to calculate the genomic kinship matrix (G-matrix) for Xinjiang Brown and Chinese Holstein cattle. Subsequently, principal component analysis (PCA) was performed using the G-matrix. The first three eigenvectors (PCA1, PCA2, and PCA3) were extracted and used as the horizontal and vertical coordinates for plotting. The contribution rates of the principal components were calculated on the basis of the percentage of eigenvalues. Finally, visualization was performed using the R language.

2.5 Multi-breed genomic evaluation using a joint reference population

2.5.1 Phenotypic data processing

The data for Xinjiang Brown cattle include production performance measurement records from 1983 to 2018 and DHI measurement records from 2010 to 2017. The data for Chinese Holstein cattle include DHI measurement records from 2001 to 2019. Milk-related traits, including 305-day milk yield (305dMY), milk fat yield (MFY), milk protein yield (MPY), and somatic cell score (SCS), were obtained through collation (Table 1). There were 7,516 and 93,717 milk trait measurements recorded for Xinjiang Brown and Chinese Holstein cattle, respectively.

The pedigree file used to analyze Xinjiang Brown cattle had 16,795 cattle, including 676 breeding bulls. Among these bulls, one had a maximum of 619 offspring, whereas 221 had only one offspring. Among the female adult cattle, 583 had only one offspring, whereas 1,623 had two or more offspring, with a maximum of 12 offspring per individual.

For the Chinese Holstein cattle, the pedigree file used for the analysis contained 6,54,390 individuals, including 11,243 breeding bulls. Among these bulls, one had a maximum of 7,884 offspring, whereas 4,695 had only one offspring. Among the female adult cattle, 1,63,781 had only one offspring, whereas 1,11,912 had two or more offspring, with a maximum of 12 offspring per individual (Table 2).

2.5.2 Genotype data

Genotype data for 403 female Xinjiang Brown cattle, 71 male Xinjiang Brown cattle, and 11 male Brown Swiss cattle was considered. In addition, 2,100 Chinese Holstein cattle were

TABLE 1 The standards for data filtering.

Breed	Character name (unit) ^a	Screening criteria
Xinjiang Brown Cattle milk	305MY(kg)	2,000–1,3000
	MFP(%)	2–7
	MPP(%)	2–7
	SCC(1,000/mL)	0–25,000
Chinese Holstein Cattle milk	305MY(kg)	4,000–15,000
	MFP(%)	2–7
	MPP(%)	2–7
	SCC(1,000/mL)	0–25,000

^a305 dMY: 305 daily milk yield; MFP: milk fat percentage; MPP: milk protein percentage; SCC: somatic cell count.

TABLE 2 Data statistics.

Breed	Number	Phenotype animals	Pedigree animals
Xinjiang Brown Cattle	7,516	2,207	16,795
Chinese Holstein Cattle	93,717	48,464	654,390
Total	1,01,233	50,671	6,71,185

TABLE 3 Gradient grouping of joint reference group.

Joint reference group	Xinjiang Brown cattle	Chinese holstein cattle
485	485	0
785	485	300
1,085	485	600
1,385	485	900
1,685	485	1,200
1985	485	1,500
2,285	485	1800
2,585	485	2,100

randomly selected (According to PCA and Admixture results, PLINK software was used to remove the chip data of Chinese Holstein cows that was inconsistent with the large population of Chinese Holstein cows).

2.5.3 Statistical analysis

In this study, the ssGBLUP method was used to construct the H-matrix based on the pedigree and genomic information from Xinjiang Brown and Chinese Holstein cattle. The two-trait model Bayesian approach was used to estimate the variance components and breeding values for each trait.

To investigate the suitable integral ratio of the Chinese Holstein cattle in the joint reference population, a random gradient grouping approach was applied. The population was gradually accumulated in increments of 300 individuals to construct the joint reference population. A control group was established by excluding the phenotypic and genomic information of the Chinese Holstein cattle (Table 3).

Because of the significant differences in milk production traits between Xinjiang Brown and Chinese Holstein cattle (Zhang et al., 2021), a dual-trait animal model was constructed. In this model, each biological trait was treated individually in the two populations, accounting for potential scale inconsistencies that may arise during breeding value estimation due to standardization across different breeds. The milk production traits (305dMY, MFY, MPY, and SCS) of Xinjiang Brown and Chinese Holstein cattle were considered to be two separate traits. A dual-trait linear model was used to estimate the variance components for milk production traits based on the genomic-pedigree combined relationship matrix, H-matrix. The model is described as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

In the formula, y_1 represents the observation value vector of a certain milk trait for each of Xinjiang Brown cattle, and y_2

represents Chinese Holstein cattle. β_1 and β_2 represent the fixed effect vectors for the same milk trait of Xinjiang Brown and Chinese Holstein cattle, respectively, including farm effect, calving year effect, calving season effect, and parity effect. The farm effect was divided into 20 levels based on the phenotypic data sources of the farms where the cattle were raised. The calving years of Xinjiang Brown cattle were divided into seven levels based on phenotypic records: 1985–1995, 1996–2000, 2001–2005, 2006–2008, 2009–2011, 2012–2014, and 2015–2018. The calving years of the Chinese Holstein cattle were divided into six levels: 2001–2005, 2006–2008, 2009–2011, 2012–2014, 2015–2018, and 2019. For the parity effect, Xinjiang Brown and Chinese Holstein cattle were classified into six levels: 1, 2, 3, 4, 5, and 6 (including those with more than six calves). The calving season effect of Xinjiang Brown cattle was divided based on the unique climatic conditions of Xinjiang. According to the method of temperature intervals, April and May were considered spring, June, July, and August were considered summer, September was considered autumn, and January, February, March, October, November, and December were considered winter. Conversely, the calving season effect of Chinese Holstein cattle was determined based on the climatic characteristics of Beijing. Following the same method of temperature intervals, March, April, and May were considered spring, June, July, and August were considered summer, September, October, and November were considered autumn, and December, January, and February were considered winter. a_1 and a_2 represent the individual additive genetic effect vectors for a certain milk trait of Xinjiang Brown and Chinese Holstein cattle, respectively. e_1 and e_2 represent the random residual effect vectors for the same milk trait of Xinjiang Brown and Chinese Holstein cattle, respectively. X_i and Z_i represent the incidence matrices for the fixed effects and individual random additive genetic effects of the i -th trait, respectively.

Assume $\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \sim N\left(0, H \otimes \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} \\ \sigma_{a_2 a_1} & \sigma_{a_2}^2 \end{bmatrix}\right)$, $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left(0, I \otimes \begin{bmatrix} \sigma_{e_1}^2 & 0 \\ 0 & \sigma_{e_2}^2 \end{bmatrix}\right)$. In the aforementioned formula, H represents the combined genomic-pedigree relationship matrix. $\sigma_{a_i}^2$ denotes the additive genetic variance for the i -th breed, while $\sigma_{a_1 a_2}$ represents the covariance between breeds. $\sigma_{e_i}^2$ stands for the residual variance of the i -th breed. Given that Xinjiang Brown and Chinese Holstein cattle are reared in separate populations, there is no residual covariance between the two groups.

The genetic variance-covariance structure of the ssGBLUP additive genetic effect model is represented by $a \sim N(0, H\sigma_a^2)$, where σ_a^2 denotes the additive genetic variance. H , the pedigree-genome relationship matrix, represents a combination of the pedigree-based additive genetic relationship matrix (A matrix) and the genome-based kinship matrix (G-matrix) (Aguilar et al., 2010; Christensen and Lund, 2010).

The formula used to compute H is as follows:

$$H = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{21} & A_{12}A_{22}^{-1}G \\ GA_{22}^{-1}A_{21} & G \end{bmatrix}$$

Subscripts 1 and 2 in A represent the non-genotyped and genotyped animals in the population, respectively. G represents the genetic relationship matrix. The calculation formula is as follows: $G = \frac{MM'}{2 \sum_{k=1}^m p_k(1-p_k)}$. M represents the

association matrix for SNP effects, where the elements $(0 - 2p_j)$, $(1 - 2p_j)$, and $(2 - 2p_j)$ represent homozygous 11, heterozygous 12 or 21, and homozygous 22 genotypes, respectively. p_j represents the minor allele frequency of the j th SNP, and m represents the number of markers. p_k denotes the allele frequency of the k th SNP. Therefore, the H^{-1} formula is given by $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} + A_{22}^{-1} \end{bmatrix}$: where A^{-1} represents the inverse matrix of all pedigree relationships, G^{-1} represents the inverse matrix of genomic kinship relationships, and A_{22}^{-1} represents the inverse matrix of the pedigree relationships for the sequenced individuals. The Bayes method was calculated using the GIBBS1F90 module in the BLUPF90 software along with the Bayes-Gibbs sampling method. In the Bayes method, the total chain length of the samples was 100,000, the burn-in chain length was 10,000, and the thinning interval was 50. The Geweke diagnostic method in POSTGIBBSF90 was used to check the convergence of the Gibbs chain (Zhang et al., 2022).

2.5.4 Calculation of heritability

The calculation formula of heritability is as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

The formula for the standard error of heritability is shown below:

$$SE^2(h^2) = \left[\frac{\sigma_a^2}{\sigma_p^2} \right] \left[\frac{Var(\sigma_a^2)}{(\sigma_a^2)^2} + \frac{Var(\sigma_p^2)}{(\sigma_p^2)^2} - \frac{Cov(\sigma_a^2, \sigma_p^2)}{\sigma_a^2 \sigma_p^2} \right]$$

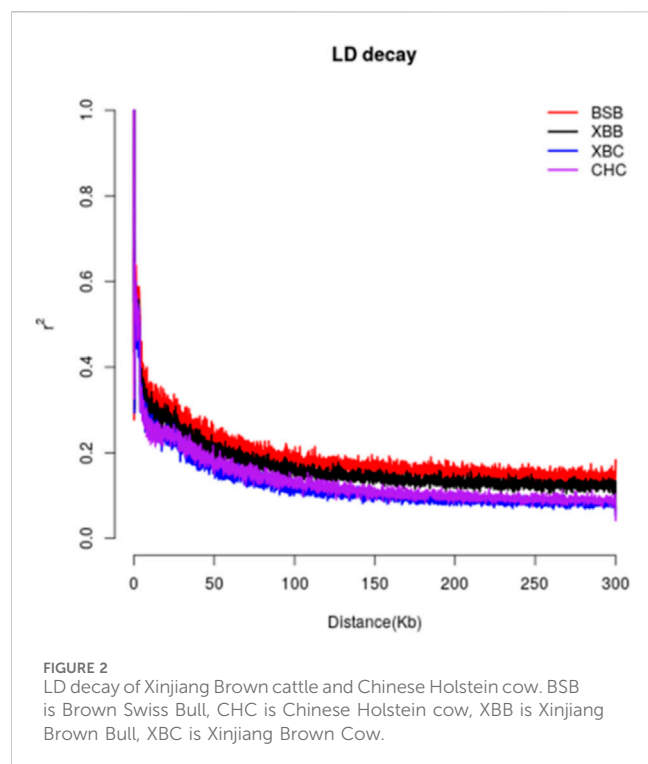
where h^2 is heritability, $SE^2(h^2)$ is the standard error of heritability, σ_a^2 is the additive genetic variance, σ_e^2 is the residual variance. σ_p^2 is the overall phenotypic variance, $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$.

2.5.5 Verification of the reliability of genomic breeding values

To verify the accuracy of the estimated genomic breeding values for the joint reference populations, 50 offspring individuals born in the past 4 years from 485 genotyped Xinjiang Brown cattle served as a validation group. Genomic predictions were performed in two groups: with and without excluding the phenotypic data of the validation group. This resulted in 16 sets of genetic parameters and genomic estimated breeding values for each milk trait. By comparing the prediction accuracy, estimation bias, and inflation coefficient of the validation group, the optimal number of joint reference populations was determined.

To calculate the prediction accuracy of the genomic estimated breeding values, the correlation coefficient between the genomic breeding values calculated with the phenotypic data of the validation group and those calculated without these data was used to measure the accuracy of estimating genomic breeding values for different joint reference populations. The formula is as follows: $R_{GEBV} = \text{Cor}(TBV^*, GEBV)$, where R^2 represents reliability or the square of accuracy.

Meanwhile, the regression of TBV^* on $GEBV$ is calculated using the formula $TBV^* = b_0 + b_1 GEBV$, where the regression coefficient b_1 is the inflation coefficient, and the intercept b_0 is the estimation bias (Legarra and Reverter, 2019).



$$b_1 = \frac{\text{cov}(\text{GEBV}, y)}{\text{var}(y)} = \frac{\text{var}(\text{GEBV})}{\text{var}(y)} < 1$$

$$b_1 = \frac{\text{Var}(\text{GEBV})}{\text{Var}(y)} = \frac{\text{Var}(\text{GEBV})\text{Var}(a)}{\text{Var}(a)\text{Var}(y)} = r_{\text{GEBV}}^2 r_y^2 \ll 1$$

In this context, b_1 represents the regression coefficient, which has the following implications: when $b_1 < 1$, GEBV is inflated, indicating that $\text{Var}(\text{GEBV})$ is too large. This means that in the genomic breeding values, the good ones are even better, and the bad ones are even worse. Conversely, when $b_1 > 1$, GEBV is deflated, suggesting that $\text{Var}(\text{GEBV})$ is too small. This indicates that the genomic breeding values are smaller than the true values and are contracted toward the middle.

3 Results

3.1 Genetic structure

3.1.1 Linkage disequilibrium analysis

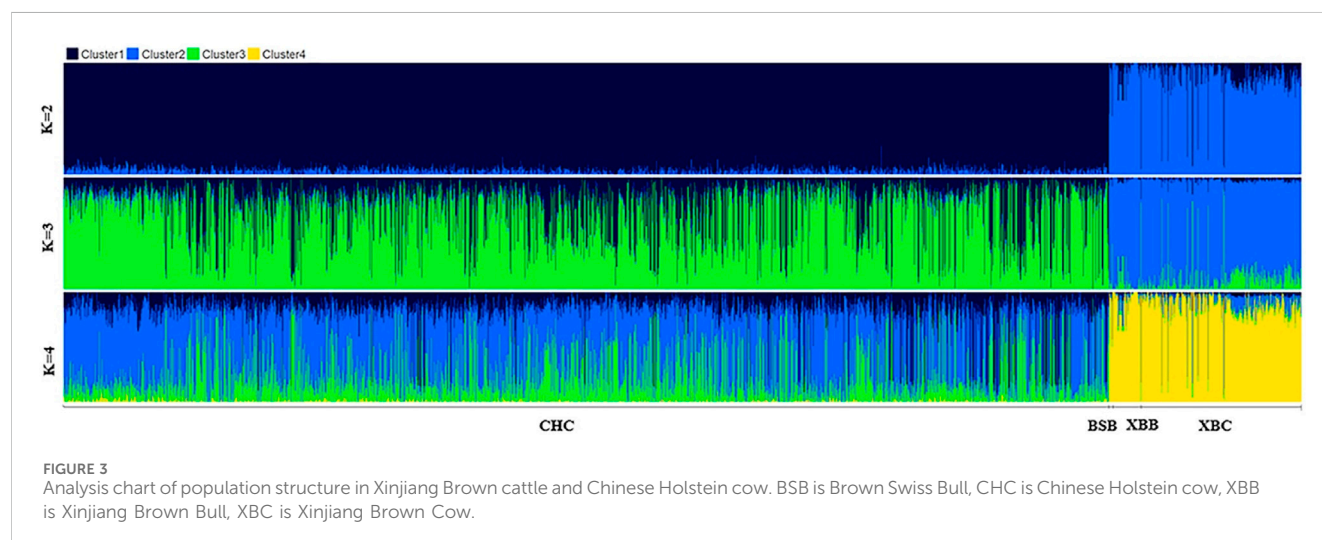
Linkage disequilibrium analysis was conducted between the two breeds by calculating the linkage disequilibrium coefficients for the two loci and plotting the LD decay graph (Figure 2). The graph shows that the average LD coefficients for Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows at a genomic distance of 50 kb were approximately 0.2, 0.25, 0.3, and 0.35, respectively, indicating a gradual increase. Noteworthy, the decay rates of the LD coefficients vary among different populations. Among the breeds, Brown Swiss bulls exhibited the slowest LD decay at 0–40 kb, whereas the Chinese Holstein cows exhibited the fastest LD decay. However, in the range of 40–300 kb, Xinjiang Brown cows exhibited the fastest LD decay, with a decay rate order of Xinjiang Brown cows > Chinese Holstein cows > Xinjiang Brown bulls > Brown Swiss bulls.

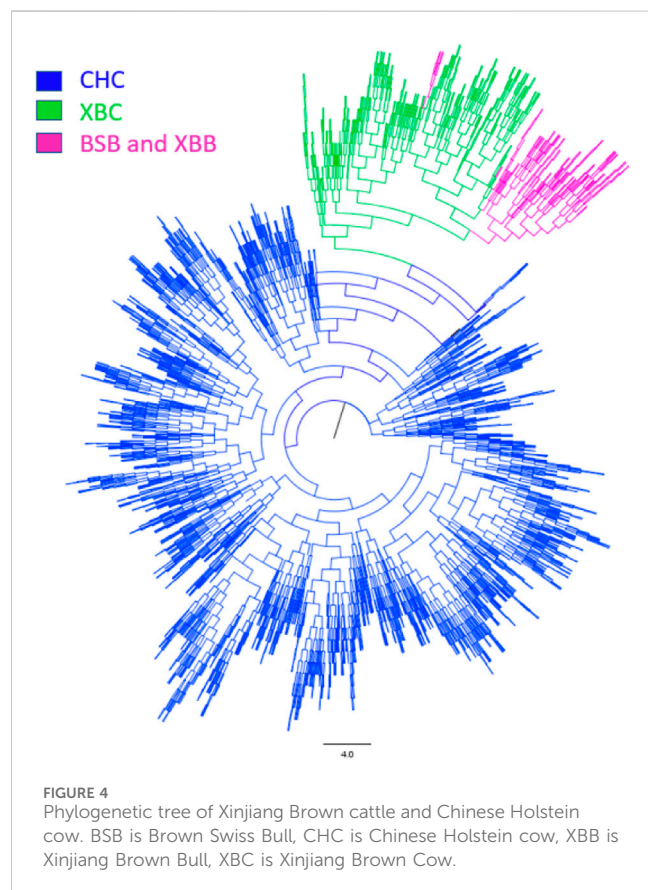
3.1.2 Population structure analysis

To further investigate the genetic components of Xinjiang Brown and Chinese Holstein cows, population structure and phylogenetic tree analyses were conducted. As shown in Figure 3, when the number of ancestral populations $K = 2$, there was a clear distinction in the genetic structure among Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows. However, the genetic structure within each group differed insignificantly. As shown in Figure 4, Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls were clustered, whereas the Chinese Holstein cows were clustered separately. In addition, Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls appear at the end of a certain branch within the Chinese Holstein population.

3.1.3 Genetic relatedness between Xinjiang Brown and Chinese Holstein cows

Using the SNP genotyping information from 403 Xinjiang Brown cows, 71 Xinjiang Brown bulls, 11 Brown Swiss bulls, and

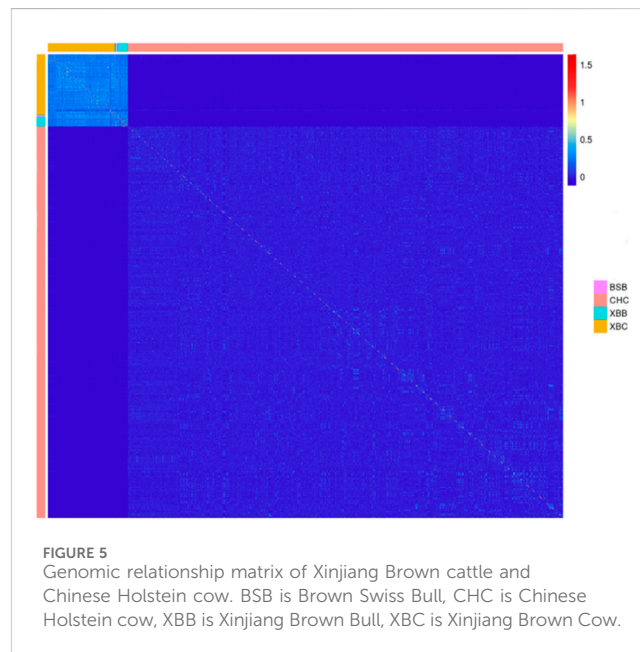




2,633 Chinese Holstein cows, a G-matrix was constructed. Figure 5 was then generated on the basis of the actual genetic relatedness among individuals in the G-matrix. Figure 5 shows that the kinship coefficients among Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls populations were approximately 0.5, which is significantly higher than those among individuals within the Chinese Holstein cow population. The kinship coefficients between Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows populations tend toward 0.

3.1.4 PCA

PCA was performed using the genomic kinship relationship matrix (G-matrix) among individuals from the two breeds (Figures 2–7). The results revealed that the first principal component (PC1, accounting for 4.75%) separated Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows into distinct groups. Specifically, Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls were closely clustered. The second principal component (PC2, accounting for 1.76%) could not distinguish between Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls; however, it separated the Chinese Holstein cows into two distinct groups. The third principal component (PC3, accounting for 1.22%) further distinguished the Chinese Holstein cows into two groups.



3.2 Multi-breed genomic evaluation using a joint reference population

3.2.1 Descriptive statistics of the dairy traits of Xinjiang brown and Chinese Holstein cattle

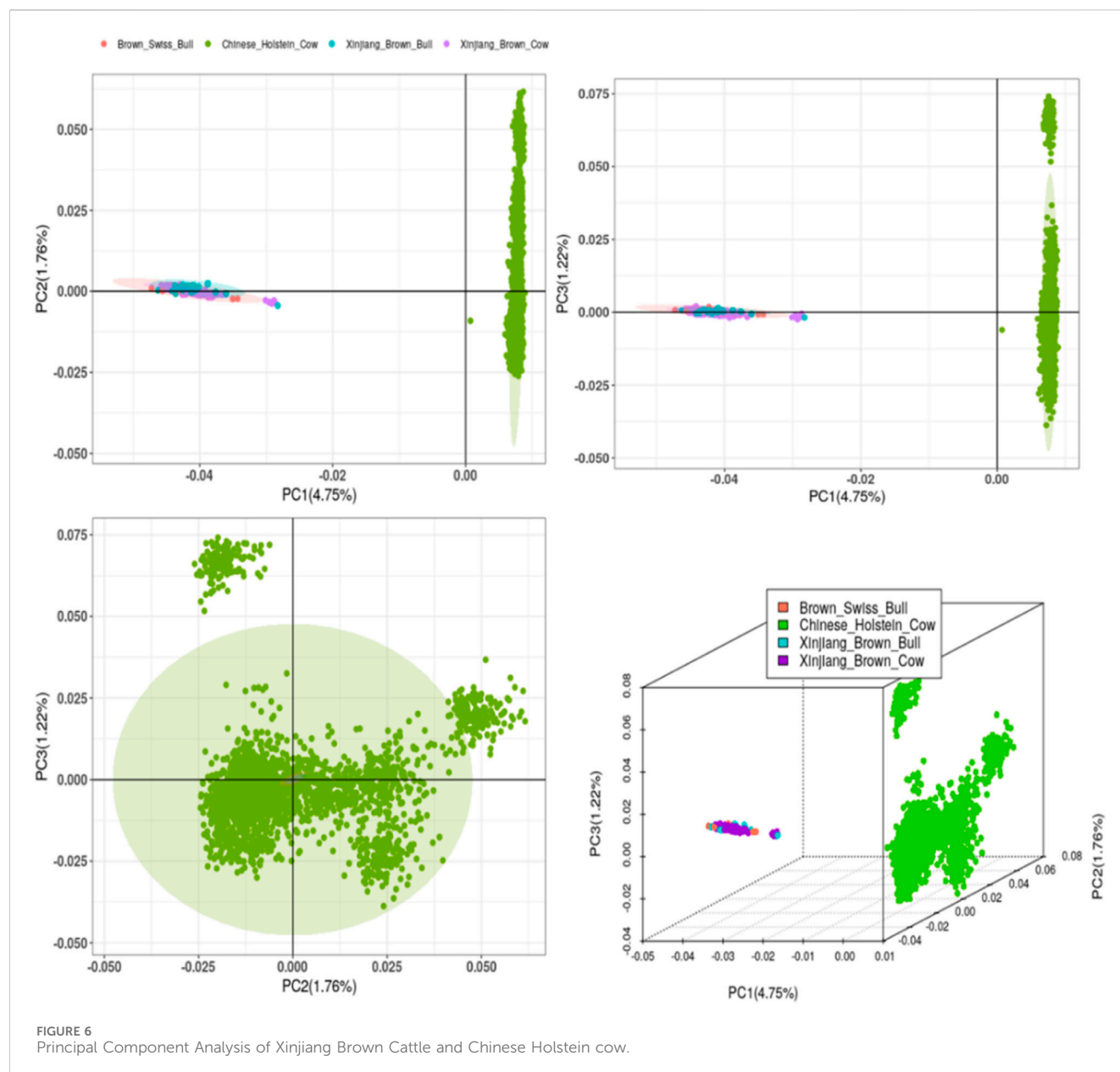
Table 4 lists the statistics, including sample size, minimum value, maximum value, mean, standard deviation, and coefficient of variation of the observed dairy traits of Xinjiang Brown and Chinese Holstein cattle. 305dMY, MFY, MPY, and SCS between Xinjiang Brown and Chinese Holstein cattle differed significantly.

3.2.2 Random grouping of the joint reference group

The chip data of 2,633 Chinese Holstein cows were screened to eliminate individuals with distant kinship within the Chinese Holstein cattle population, leaving 2,271 genotyped Chinese Holstein cows. Among these, 2,100 cows were randomly selected as the total population of Chinese Holstein cows to be included in the joint reference group. Random equal-sized groupings were then performed on 2,100 genotyped Chinese Holstein cows, resulting in subsets with different numbers of cows. As shown in Figure 7, the distribution of the subsets in the total population was relatively scattered for the Chinese Holstein cows added to the joint reference group.

3.2.3 Estimation of genetic parameters of dairy traits in the joint reference group

As shown in Table 5, before the inclusion of the Chinese Holstein population, the heritability of 305dMY was 0.204 without excluding the phenotypic data of 50 genotype Xinjiang Brown cows, which decreased to 0.203 after data exclusion. When varying numbers of Chinese Holstein cows were incorporated into the joint reference group, the heritability of



305dMY in Xinjiang Brown cows was 0.137–0.249 without excluding the phenotypic data of the 50 genotyped cows. However, after excluding these data, the heritability of 305dMY in Xinjiang Brown cows was adjusted to 0.169–0.254.

As shown in Table 6, without the inclusion of Chinese Holstein cows, the heritability of MFY, was 0.07 when the phenotypic data of 50 genotype Xinjiang Brown cows were included, and it was 0.073 when the phenotypic data were excluded. When different numbers of Chinese Holstein cows were added to the reference population, the MFY, heritability of Xinjiang Brown cows was 0.073–0.086 when the phenotypic data of 50 genotype Xinjiang Brown cows were included, and it was 0.057–0.085 when the phenotypic data were excluded.

As shown in Table 7, without the inclusion of the Chinese Holstein cows, the heritability of MPY, was 0.143 when the

phenotypic data of 50 genotyped Xinjiang Brown cows were included, and it was 0.145 when the phenotypic data were excluded. When different numbers of Chinese Holstein cows were added to the reference population, the MPY, heritability of Xinjiang Brown cows was 0.123–0.158 when the phenotypic data of the 50 genotyped Xinjiang Brown cows were included, and it was 0.0142–0.174 when the phenotypic data were excluded.

Table 8 shows that without the inclusion of the Chinese Holstein cow population, the heritability of SCS was 0.042 when the phenotypic data of the 50 genotyped Xinjiang Brown cows were included and 0.043 when the phenotypic data were excluded. After adding different numbers of the Chinese Holstein cows to the joint reference population, the SCS heritability for Xinjiang Brown cows was 0.02–0.062 when the phenotypic data of the 50 genotyped Xinjiang Brown cows were included. When the phenotypic data were excluded, the SCS heritability for Xinjiang Brown cows was 0.015–0.081.

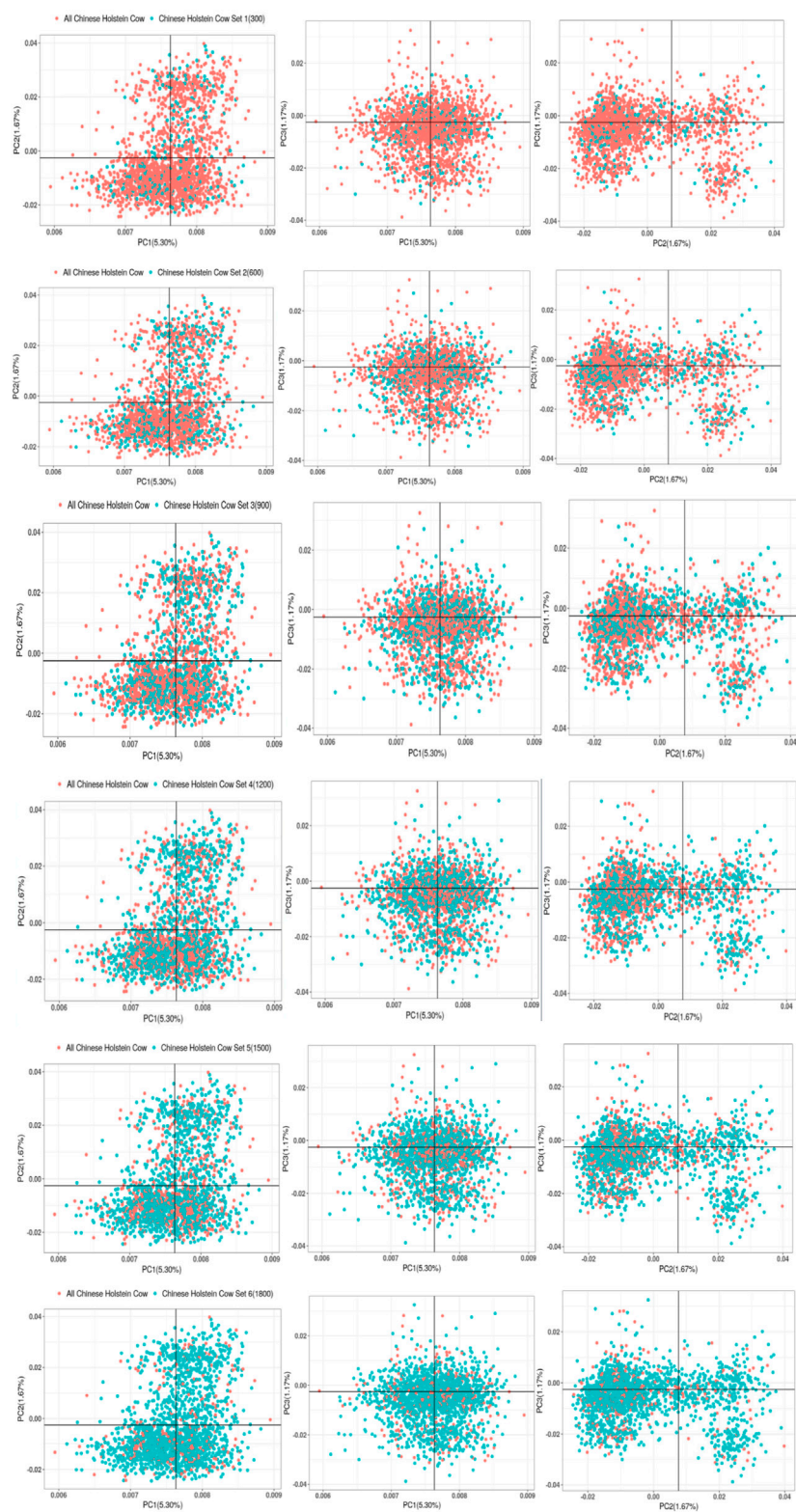


FIGURE 7
Gradient grouping of PCA analysis in joint reference group.

TABLE 4 Description of milk traits in Xinjiang Brown Cattle and Chinese Holstein Cattle.

Breed	Trait ^a	Number	Minimum	Maximum	Average	SD	CV(%)
Xinjiang Brown	305dMY/kg	7,515	814	8,444	4126.49	1405.71	34.07
	MFY/kg	2,655	21.6	431.55	168.53	68.29	40.52
	MPY/kg	2,655	20.3	302.72	143.71	51.42	35.78
	SCS	2,655	−2.05	10.95	4.98	2.16	43.37
Chinese Holstein	305dMY/kg	89,350	4001	15,000	10116.68	2045.07	20.21
	MFY/kg	89,350	86.30	1019.21	398.52	117.43	29.47
	MPY/kg	89,350	87.72	951.76	329.92	74.40	22.55
	SCS	89,350	−3.65	9.65	3.25	1.81	55.66

^a305 dMY: 305 daily milk yield; MFY: milk fat yield; MPY: milk protein yield; SCS: somatic cell score. SD: Standard deviation.

3.2.4 Verification of the genomic breeding value reliability

As shown in Table 9, when different numbers of Chinese Holstein cows were added to the joint reference population, the reliability of the total population genomic breeding values for 305dMY of Xinjiang Brown cows was 0.142–0.340, with a regression coefficient of 0.129–0.312. The reliability of the genomic breeding values for the validation population was −0.033–0.087, with a regression coefficient of −0.064–0.056.

As shown in Table 10, when different numbers of the Chinese Holstein cows were added to the joint reference population, the reliability of the total population genomic breeding values for MFY, of Xinjiang Brown cows was 0.263–0.424, with a regression coefficient of 0.296–0.437. The reliability of the genomic breeding values for the validation population was −0.149–0.138, with a regression coefficient of −0.192–0.137.

As shown in Table 11, when different numbers of Chinese Holstein cows were added to the joint reference population, the reliability of the total population genomic breeding values for MPY, of Xinjiang Brown cows was 0.28–0.465, with a regression coefficient of 0.277–0.504. The reliability of the genomic breeding values for the validation population was −0.259–0.203, with a regression coefficient of −0.213–0.032.

As shown in Table 12, when different numbers of Chinese Holstein cows were added to the joint reference population, the reliability of the total population genomic breeding values for SCS, of Xinjiang Brown cows was 0.190–0.448, with a regression coefficient of 0.207–0.502. The reliability of the genomic breeding values for the validation population was −0.145–0.062, with a regression coefficient of −0.223–0.075.

4 Discussion

4.1 Genetic structure analysis

LD is a metric that quantifies whether genotype variations in two SNP markers are relatively consistent and whether they are correlated (Park, 2012). If two loci with adjacent alleles are correlated, certain genotypes tend to be co-inherited, resulting in

a higher frequency of certain haplotypes than expected. This pattern can be visually represented using an LD decay plot (Amaral et al., 2008). Because of the correlated inheritance of the two loci, the decay rate of the LD coefficient decreases with increasing generations and recombination events. Genetic background also influences it. Domestication selection reduces genetic diversity within a population, reinforcing the correlation or linkage between SNP loci (Farnir et al., 2000). Consequently, populations with higher degrees of domestication exhibit stronger selection intensities (Odani et al., 2006), resulting in slower LD decay rates. The higher selection intensity in breeding bulls likely reduced the effective population size, thereby affecting LD in these groups. Meanwhile, the LD decay patterns in Xinjiang Brown and Chinese Holstein cows exhibit potential similarities, favoring the construction of a combined reference population. Overall, the LD decayed fastest in Xinjiang Brown cows, indicating lower levels of selection than the other three groups. This suggests a high level of genetic diversity in Xinjiang Brown cows, which harbor rich genetic resources with potential for development and use. These findings provide a scientific basis for the conservation, exploitation, and use of genetic diversity in Xinjiang Brown cattle.

Genetic structure analysis can elucidate phylogenetic relationships and genetic distances among different populations (Whelan and Goldman, 2001). Under the influence of natural and artificial selection, populations exhibiting pronounced genetic differences are evident. When $K = 2$, there is a distinct genetic structure differentiation among Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows. This is related to the breeding strategies employed during the intense selection process of Xinjiang Brown and Chinese Holstein cattle, where most female progenitors in the early stages of population breeding originated from local Chinese yellow cattle (Liu, 2013).

The breeding of Xinjiang Brown and Chinese Holstein cattle involves the introduction of foreign breeds for crossbreeding to improve and enhance the local yellow cattle population in China. Subsequently, through crossbreeding fixation and selective breeding, these breeds have been further developed and stabilized. The genetic background of Xinjiang Brown cattle is traceable to the original crossbreeding improvement in 1951, when the maternal breed was Kazakh cattle (Ma, 2015). Conversely, the genetic background of the

TABLE 5 Genetic parameter estimation of 305dMY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Breed	All data			All data-50		
				σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)	σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)
485	485	0	Xinjiang Brown	235,810	925,050	0.204	237,340	931,960	0.203
				(23,428)	(19,516)	(0.018)	(23,720)	(19,687)	(0.018)
785	485	300	Xinjiang Brown	395,790	2,495,000	0.137	605,190	2,378,800	0.203
				(256,700)	(292,170)	(0.081)	(236,900)	(272,020)	(0.072)
785	485	300	Chinese Holstein	245,370	923,400	0.21	245,710	930,680	0.209
				(24,660)	(19,243)	(0.019)	(24,693)	(19,285)	(0.019)
1,085	485	600	Xinjiang Brown	768,210	2,324,100	0.249	785,140	2,313,700	0.254
				(205,040)	(181,580)	(0.057)	(192,460)	(180,370)	(0.054)
1,085	485	600	Chinese Holstein	246,200	922,650	0.211	246,250	930,390	0.21
				(24,449)	(18,966)	(0.019)	(24,649)	(19,962)	(0.019)
1,385	485	900	Xinjiang Brown	635,210	2,316,300	0.216	544,100	2,363,400	0.188
				(125,820)	(129,040)	(0.038)	(122,830)	(136,000)	(0.039)
1,385	485	900	Chinese Holstein	243,990	923,970	0.209	249,440	929,320	0.212
				(24,038)	(19,290)	(0.018)	(24,562)	(19,131)	(0.018)
1,685	485	1,200	Xinjiang Brown	591,430	2,562,300	0.188	523,940	2,586,200	0.169
				(130,800)	(123,390)	(0.037)	(75,741)	(110,240)	(0.023)
1,685	485	1,200	Chinese Holstein	243,880	923,950	0.209	253,520	926,610	0.215
				(24,708)	(19,457)	(0.019)	(23,774)	(19,612)	(0.018)
1985	485	1,500	Xinjiang Brown	610,630	2,556,100	0.193	582,790	2,571,100	0.185
				(107,850)	(101,890)	(0.031)	(119,430)	(105,660)	(0.035)
1985	485	1,500	Chinese Holstein	242,740	923,660	0.209	246,930	929,650	0.21
				(24,064)	(19,028)	(0.018)	(24,018)	(19,202)	(0.018)
2,285	485	1800	Xinjiang Brown	629,300	2,573,000	0.197	637,850	2,569,600	0.199
				(97,901)	(92,596)	(0.028)	(102,810)	(94,655)	(0.029)
2,285	485	1800	Chinese Holstein	244,750	923,220	0.21	246,040	930,360	0.21
				(23,787)	(18,710)	(0.018)	(25,235)	(19,941)	(0.019)
2,585	485	2,100	Xinjiang Brown	654,250	2,513,800	0.207	642,470	2,520,000	0.204
				(91,086)	(85,262)	(0.026)	(91,676)	(83,306)	(0.026)
2,585	485	2,100	Chinese Holstein	242,120	924,250	0.208	247,120	929,710	0.21
				(23,562)	(19,207)	(0.018)	(25,612)	(19,792)	(0.019)

Note: "All data" refers to the calculation results without excluding the phenotypic data of the 50 validation animals; "All data-50" refers to the calculation results after excluding the phenotypic data of the 50 validation animals. σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h^2 = heritability; SE, standard error.

Chinese Holstein cattle dates back to 1840. From 1840 to 1948, the Chinese Holstein cattle underwent more than a century of introduction and early stages of crossbreeding improvement. During this period, China initially introduced various dairy breeds, including the Holstein cattle, Jersey cattle, Ayrshire cattle, Brown Swiss cattle, and Shorthorn cattle (Liu, 2013). The distinct genetic structure observed among Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows

indicates a significant genetic distance between these two major groups. This significantly affects the genetic structure of multi-breed joint reference genomes established later, subsequently affecting the accuracy of genomic predictions.

Figure 5 shows that the coefficient of kinship among individuals within the populations of Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls was relatively high. However, the coefficient of kinship between these groups and the Chinese

TABLE 6 Genetic parameter estimation of MFY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Breed	All data			All data-50		
				σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)	σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)
485	485	0	Xinjiang Brown	211.01	2843.6	0.07	225.03	2879.9	0.073
				(64.235)	(94.068)	(0.021)	(67.683)	(97.71)	(0.022)
785	485	300	Xinjiang Brown	260.07	2820.2	0.085	261.65	2860.5	0.084
				(70.93)	(92.423)	(0.023)	(73.108)	(97.286)	(0.023)
785	485	300	Chinese Holstein	667.02	10,151	0.062	1,091	9870.2	0.1
				(628.76)	(1056.8)	(0.056)	(663.79)	(1014.2)	(0.057)
1,085	485	600	Xinjiang Brown	261.42	2821.7	0.085	260.63	2863.3	0.084
				(72.708)	(92.754)	(0.023)	(68.761)	(96.963)	(0.022)
1,085	485	600	Chinese Holstein	1,248	9922.6	0.112	1594.4	9682.6	0.142
				(522.17)	(703.23)	(0.045)	(544.89)	(678.09)	(0.046)
1,385	485	900	Xinjiang Brown	262.8	2819.4	0.086	251.19	2865.4	0.081
				(66.884)	(93.427)	(0.021)	(77.111)	(97.509)	(0.024)
1,385	485	900	Chinese Holstein	1015.3	10,726	0.087	734.34	10,921	0.064
				(414.44)	(553.17)	(0.034)	(511.15)	(601.11)	(0.043)
1,685	485	1,200	Xinjiang Brown	256.98	2821.5	0.084	263.17	2860.9	0.085
				(67.492)	(94.732)	(0.022)	(80.748)	(97.523)	(0.025)
1,685	485	1,200	Chinese Holstein	715.82	10,683	0.063	838.62	10,581	0.074
				(372.83)	(497.08)	(0.032)	(339.32)	(467.88)	(0.029)
1985	485	1,500	Xinjiang Brown	253.46	2820.4	0.083	249.42	2866.8	0.081
				(58.914)	(91.339)	(0.019)	(75.035)	(96.106)	(0.024)
1985	485	1,500	Chinese Holstein	629.08	10,465	0.057	785.9	10,349	0.071
				(317.84)	(417.27)	(0.028)	(285.16)	(399.72)	(0.025)
2,285	485	1800	Xinjiang Brown	266.36	2818.9	0.087	242.33	2871.7	0.078
				(73.755)	(97.11)	(0.023)	(61.402)	(94.469)	(0.02)
2,285	485	1800	Chinese Holstein	717.61	10,529	0.064	754.32	10,503	0.068
				(239.88)	(360.06)	(0.021)	(266.09)	(364.18)	(0.024)
2,585	485	2,100	Xinjiang Brown	223.39	2843.5	0.073	174.8	2,917	0.057
				(67.122)	(91.748)	(0.022)	(74.579)	(101.76)	(0.024)
2,585	485	2,100	Chinese Holstein	1037.2	10,437	0.091	1139.2	10,376	0.099
				(256.92)	(336.28)	(0.022)	(249.42)	(329.61)	(0.021)

Note: "All data" refers to the calculation results without excluding the phenotypic data of the 50 validation animals; "All data-50" refers to the calculation results after excluding the phenotypic data of the 50 validation animals. σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h^2 = heritability; SE, standard error.

Holstein cow population tended to be close to 0. This result suggests that when conducting multi-breed genomic selection, it is not advisable to use only one breed as the reference population and the other as the selection population. Instead, it is necessary to fully consider the relationship between the selection and reference populations. When the kinship between the two breeds involved in multi-breed genomic selection is weak or extremely weak, it is necessary to include a certain number of individuals from the same

breed in the reference population to ensure high reliability in the estimation of genomic breeding values.

The PCA results identified two major clusters: one containing Xinjiang Brown cows, Xinjiang Brown bulls, Brown Swiss bulls, and Chinese Holstein cows, whereas the other mainly contained Chinese Holstein cows. This suggests a relatively distant genetic relationship between these two clusters. Clustering of Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls is highly concentrated,

TABLE 7 Genetic parameter estimation of MPY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Breed	All data			All data-50		
				σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)	σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)
485	485	0	Xinjiang Brown	238.44	1429.4	0.143	246.05	1455.9	0.145
				(46.431)	(48.997)	(0.026)	(47.902)	(50.543)	(0.027)
785	485	300	Xinjiang Brown	253.09	1423.1	0.151	267.32	1446.7	0.156
				(50.485)	(49.704)	(0.028)	(48.616)	(48.525)	(0.027)
785	485	300	Chinese Holstein	463.33	3956.7	0.105	451.38	3955.5	0.103
				(477.27)	(464.24)	(0.1)	(382.5)	(461.92)	(0.081)
1,085	485	600	Xinjiang Brown	261.17	1419.4	0.156	264.41	1450.2	0.155
				(48.804)	(47.058)	(0.027)	(51.41)	(49.689)	(0.028)
1,085	485	600	Chinese Holstein	786.7	3971.3	0.166	828.05	3935.7	0.174
				(310.95)	(308.68)	(0.061)	(280.6)	(277.55)	(0.053)
1,385	485	900	Xinjiang Brown	264.73	1419.4	0.158	265.12	1450.5	0.155
				(51.675)	(48.283)	(0.029)	(47.87)	(49.206)	(0.026)
1,385	485	900	Chinese Holstein	544.43	3886.4	0.123	593.77	3858.1	0.134
				(188.04)	(222.02)	(0.041)	(172.04)	(204.45)	(0.036)
1,685	485	1,200	Xinjiang Brown	255.49	1423.7	0.153	265.08	1447.8	0.155
				(49.583)	(48.699)	(0.028)	(50.392)	(52.497)	(0.028)
1,685	485	1,200	Chinese Holstein	576.71	3958.4	0.128	560.76	3962.6	0.124
				(140.64)	(175.24)	(0.03)	(149.43)	(183.38)	(0.032)
1985	485	1,500	Xinjiang Brown	257.41	1422.2	0.154	250.87	1457.3	0.147
				(48.761)	(48.203)	(0.027)	(52.614)	(52.575)	(0.029)
1985	485	1,500	Chinese Holstein	563.84	3980.9	0.125	576.8	3988.6	0.127
				(136.03)	(154.87)	(0.029)	(118.89)	(155.55)	(0.025)
2,285	485	1800	Xinjiang Brown	249.04	1424.2	0.149	261.17	1451.4	0.153
				(48.875)	(48.052)	(0.028)	(48.144)	(49.9)	(0.026)
2,285	485	1800	Chinese Holstein	653.65	3916.5	0.144	645.01	3918.3	0.142
				(115.42)	(138.07)	(0.024)	(128.01)	(138.72)	(0.027)
2,585	485	2,100	Xinjiang Brown	246.36	1427.5	0.148	243.18	1458.6	0.143
				(41.863)	(47.265)	(0.024)	(52.403)	(51)	(0.029)
2,585	485	2,100	Chinese Holstein	7,112.74	3795.6	0.158	717.43	3796.2	0.159
				(118.26)	(126.31)	(0.025)	(116.58)	(119.39)	(0.024)

Note: "All data" refers to the calculation results without excluding the phenotypic data of the 50 validation animals; "All data-50" refers to the calculation results after excluding the phenotypic data of the 50 validation animals. σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h^2 = heritability; SE, standard error.

indicating a close genetic distance among these groups. When considering only the clustering of Chinese Holstein cows, a small subset can be distinguished from the larger population, which exhibits a distant genetic relationship with Xinjiang Brown cows, Xinjiang Brown bulls, and Brown Swiss bulls and with most Chinese Holstein cows. In the subsequent multi-breed genetic evaluations, it is recommended to exclude chip and phenotypic data from this subset of Chinese Holstein cows and their respective farms. This approach can help reduce the interference of genetic structure differences while

estimating SNP effects. These findings suggest a low genetic linkage between Xinjiang Brown and Chinese Holstein cattle.

4.2 Genetic parameter analysis of the joint reference population

The number of genotyped Chinese Holstein cattle in the joint reference population significantly affects the estimation of variance

TABLE 8 Genetic parameter estimation of SCS in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Breed	All data			All data-50		
				σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)	σ_a^2 (SE)	σ_e^2 (SE)	h^2 (SE)
485	485	0	Xinjiang Brown	0.174	4.043	0.042	0.18	4.015	0.043
				(0.073)	(0.127)	(0.017)	(0.079)	(0.131)	(0.019)
785	485	300	Xinjiang Brown	0.143	2.721	0.05	0.235	2.687	0.081
				(0.073)	(0.124)	(0.045)	(0.204)	(0.271)	(0.065)
785	485	300	Chinese Holstein	0.216	4.026	0.051	0.184	4.029	0.044
				(0.077)	(0.124)	(0.018)	(0.089)	(0.131)	(0.021)
1,085	485	600	Xinjiang Brown	0.159	2.718	0.056	0.106	2.76	0.037
				(0.107)	(0.18)	(0.037)	(0.113)	(0.181)	(0.039)
1,085	485	600	Chinese Holstein	0.213	4.033	0.051	0.234	3.99	0.056
				(0.802)	(0.129)	(0.019)	(0.855)	(0.127)	(0.02)
1,385	485	900	Xinjiang Brown	0.177	2.68	0.062	0.125	2.711	0.044
				(0.088)	(0.133)	(0.03)	(0.083)	(0.138)	(0.029)
1,385	485	900	Chinese Holstein	0.199	4.039	0.047	0.232	3.992	0.055
				(0.073)	(0.123)	(0.017)	(0.081)	(0.127)	(0.019)
1,685	485	1,200	Xinjiang Brown	0.139	2.891	0.046	0.134	2.896	0.045
				(0.039)	(0.112)	(0.013)	(0.055)	(0.114)	(0.018)
1,685	485	1,200	Chinese Holstein	0.218	4.027	0.052	0.2	4.01	0.048
				(0.058)	(0.123)	(0.014)	(0.071)	(0.127)	(0.017)
1985	485	1,500	Xinjiang Brown	0.09	2.899	0.03	0.918	2.896	0.031
				(0.046)	(0.101)	(0.016)	(0.463)	(0.104)	(0.016)
1985	485	1,500	Chinese Holstein	0.226	4.021	0.054	0.229	3.988	0.055
				(0.078)	(0.124)	(0.018)	(0.077)	(0.125)	(0.018)
2,285	485	1800	Xinjiang Brown	0.058	2.923	0.02	0.074	2.913	0.025
				(0.043)	(0.095)	(0.015)	(0.042)	(0.091)	(0.014)
2,285	485	1800	Chinese Holstein	0.208	4.03	0.049	0.201	4.011	0.048
				(0.069)	(0.122)	(0.016)	(0.089)	(0.13)	(0.021)
2,585	485	2,100	Xinjiang Brown	0.057	2.863	0.02	0.044	2.868	0.015
				(0.044)	(0.082)	(0.015)	(0.032)	(0.084)	(0.011)
2,585	485	2,100	Chinese Holstein	0.218	4.024	0.052	0.197	4.014	0.047
				(0.082)	(0.129)	(0.019)	(0.099)	(0.133)	(0.023)

Note: "All data" refers to the calculation results without excluding the phenotypic data of the 50 validation animals; "All data-50" refers to the calculation results after excluding the phenotypic data of the 50 validation animals. σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h^2 = heritability; SE, standard error.

components. Therefore, it is necessary to consider the number of genotyped animals from different populations in the joint reference population during multi-breed genetic evaluations. The most important factors that affect the reliability of genomic breeding value estimation are the proportion of genetic variance explained by SNPs and trait heritability (Steyn et al., 2019; van Grevenhof et al., 2019). These genetic parameters are directly related to the size and structure of the training population as well as the range, quality, and quantity of phenotypic and genomic information available for

individuals in the training population. To ensure accurate genomic breeding value estimation, it is important to minimize the relationship between genotyped individuals within the training population and maximize the relationship between the training and prediction populations. Because ssGBLUP generates genomic breeding values for cows, it is particularly useful for cows with only parental average information. The single-step genomic evaluation combines information from all countries, considering potential duplicate counting of the same information, thereby

TABLE 9 Genetic parameter estimation of 305dMY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Total population			Validation population		
			R^2_{GEBV}	b_0	b_1	R^2_{GEBV}	b_0	b_1
485	485	0	0.172	-66.82	0.171	0.225	-101.864	0.42
785	485	300	0.272	-27.443	0.271	-0.041	-36.453	-0.034
1,085	485	600	0.196	-25.215	0.201	0.087	-38.312	0.056
1,385	485	900	0.142	-25.038	0.129	0.04	-38.802	0.027
1,685	485	1,200	0.340	-12.308	0.312	-0.061	-64.426	-0.037
1985	485	1,500	0.245	-12.074	0.242	-0.083	-72.544	-0.064
2,285	485	1800	0.302	-14.355	0.291	-0.033	-71.878	-0.024
2,585	485	2,100	0.205	-18.233	0.257	-0.147	-80.443	-0.13

TABLE 10 Genetic parameter estimation of MFY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Total population			Validation population		
			R^2_{GEBV}	b_0	b_1	R^2_{GEBV}	b_0	b_1
485	485	0	0.300	-1.492	0.296	-0.141	-0.776	-0.083
785	485	300	0.414	-0.332	0.437	0.035	-2.114	0.038
1,085	485	600	0.404	-0.236	0.413	0.138	-2.734	0.137
1,385	485	900	0.374	-0.189	0.41	-0.065	-3.829	-0.061
1,685	485	1,200	0.379	-0.271	0.389	0.038	-3.699	0.033
1985	485	1,500	0.382	-0.227	0.374	-0.132	-4.211	-0.101
2,285	485	1800	0.263	-0.303	0.297	-0.149	-4.21	-0.192
2,585	485	2,100	0.424	-0.253	0.393	-0.03	-4.451	-0.032

TABLE 11 Genetic parameter estimation of MPY in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Total population			Validation population		
			R^2_{GEBV}	b_0	b_1	R^2_{GEBV}	b_0	b_1
485	485	0	0.280	-2.234	0.277	0.203	-1.654	0.136
785	485	300	0.465	-0.307	0.468	-0.066	-4.338	-0.048
1,085	485	600	0.444	-0.185	0.458	-0.158	-5.416	-0.116
1,385	485	900	0.421	-0.17	0.418	0.04	-5.773	0.032
1,685	485	1,200	0.386	-0.28	0.356	-0.091	-6.293	-0.079
1985	485	1,500	0.399	-0.243	0.401	-0.259	-6.967	-0.213
2,285	485	1800	0.397	-0.343	0.504	-0.165	-5.99	-0.162
2,585	485	2,100	0.371	-0.234	0.3	-0.222	-6.578	-0.176

ensuring more accurate estimation of genomic breeding values. Adding carefully selected cows to the training population can expand the population, improve its structure and relationship with the prediction population, and reduce selection bias. However, in this study, a large proportion of the chip data came

from cows, with relatively fewer bulls for validation, probably being a reason for the poor prediction performance.

The ssGBLUP method can simultaneously analyze phenotypic, genomic, and pedigree information from both genotype and non-genotype animals by integrating external information. This method

TABLE 12 Genetic parameter estimation of SCS in joint reference population based on ssGBLUP.

Joint reference population	Xinjiang Brown	Chinese holstein	Total population			Validation population		
			R^2_{GEBV}	b_0	b_1	R^2_{GEBV}	b_0	b_1
485	485	0	0.245	0.011	0.249	0.04	0.007	0.025
785	485	300	0.255	0.01	0.295	−0.145	−0.066	−0.223
1,085	485	600	0.234	0.008	0.232	0.002	−0.071	0.002
1,385	485	900	0.221	0.007	0.207	0.062	−0.084	0.075
1,685	485	1,200	0.448	0.002	0.502	−0.4	−0.084	−0.501
1985	485	1,500	0.248	0.006	0.255	−0.117	−0.089	−0.178
2,285	485	1800	0.254	0.002	0.29	0.005	−0.073	0.006
2,585	485	2,100	0.190	0.003	0.239	−0.093	−0.089	−0.138

is particularly convenient when using foreign paternal genetic material. For example, during the Interbull evaluation of Brown Swiss bulls, a country can obtain genomic information from multiple countries and MACE information. Combining external MACE information with ssGBLUP can complement paternal information from different countries and provide pseudophenotypic information for foreign paternal lines with no or few offsprings. This research result suggests that we can conduct cross-country genetic evaluations with the Brown Swiss bull origin introduced during the breeding of Xinjiang Brown cattle, which could improve the reliability of genomic predictions for Xinjiang Brown cattle.

4.3 Reliability of genetic evaluation in joint reference populations

In theory, a model that assumes the closest distribution of SNP effects to their true distribution can achieve the highest reliability in genomic prediction. The GBLUP model assumes that all SNP effects follow the same normal distribution and compresses the effects of all SNPs to the same degree because different models have different assumptions about the distribution of SNP effects (Villarh  n  dez et al., 2021). Several methods have been proposed to improve the accuracy of genomic prediction in small populations of dairy cattle (Marjanovic et al., 2021), and one effective approach is to use joint reference populations by combining reference data from different populations (Steyn et al., 2019; van Grevenhof et al., 2019). This method has reported significant benefits in genomic prediction for North American Holstein, European Holstein, Chinese Holstein, and Brown Swiss populations (Vanderick et al., 2017; van den berg et al., 2016). However, the accuracy of genomic prediction is dependent on the relationship between candidate and reference animals, requiring the reference population to be sufficiently close to the target population (Xu et al., 2019). Therefore, for multi-breed joint genetic evaluation between Xinjiang Brown and Chinese Holstein cattle, considering only the added number of animals is insufficient. Further in-depth analysis of important influencing factors, including assumptions about SNP effects (van den berg et al., 2019) and the weights of the A- and G-matrices in the

H-matrix (Karaman et al., 2018; Botelho et al., 2021), is required to improve the accuracy and unbiasedness of predictions (Botelho et al., 2021).

Including cows in the genotyping reference population is necessary because of the limited number of cows with reliable phenotypic information available for predicting offspring traits. To improve the genomic breeding values of the population, it is necessary to include a certain number of validated bulls with reliable phenotypic information in the reference population (Vanraden et al., 2020). Previous studies have reported that the inclusion of cows in the validated bull reference population can improve the accuracy of genomic prediction (Ding et al., 2013). Although the phenotypic information for cows is less accurate than that for bulls with offspring validation, additional information can still be significant (Cole et al., 2021), considering the large number of cows available as reference animals.

5 Conclusion

The genetic structure of mature Xinjiang Brown and Chinese Holstein cows is different, and the individual kinship between these two populations is relatively distant. This increases the impact of genetic structure and kinship on the reliability of genomic breeding value estimation. Through comparisons of parameters, including heritability, breeding value reliability, and unbiasedness, it was initially determined that including 600 and 900 Chinese Holstein cows in the joint reference population positively impacted the genomic prediction of Xinjiang Brown cattle to some extent. In multi-breed genome selection, it is necessary to pre-evaluate the genetic structure and genetic relationship of the population. It is feasible to combine the Chinese Holstein cattle population into the Xinjiang brown cattle population to form a joint reference group for cross-breed genetic assessment. It can provide theoretical guidance for applied genomic genetic assessment and multi-breed genomic genetic assessment of Xinjiang brown cattle, and also provide reference for genome selection of other dual-use cattle and small population breeds.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal studies were approved by Experimental animal Welfare Ethics Committee, Xinjiang Agricultural University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

MZ: Investigation, Writing—original draft, Writing—review and editing. LX: Data curation, Writing—review and editing. HnL: Writing—original draft. HnL: Methodology, Writing—original draft. JZ: Formal Analysis, Investigation, Writing—review and editing. DW: Investigation, Writing—review and editing. XZ: Investigation, Writing—review and editing. XH: Conceptualization, Resources, Supervision, Writing—review and editing. YW: Conceptualization, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work

was funded by the Xinjiang Uygur Autonomous Region Science and technology Aid Xinjiang Project (Grant No. 2018E02052), The Xinjiang Uygur Autonomous Region University Scientific Research Project (Grant No. XJEDU 2017I005) and National Modern Agriculture Industrial System Project (Grant No. CARS-36).

Acknowledgments

We greatly thank the Xinjiang Tianshan Animal Husbandry Bio-Engineering Co., Ltd., Xinjiang Uygur Autonomous Region local state-owned Urumqi cattle farm, Yili Xinjiang Brown cattle farm, Tacheng Agriculture and Animal Husbandry Technology Co., Ltd., China for data and samples collection.

Conflict of interest

Author JZ was employed by Shijiazhuang Molbreeding Biotechnology Co., 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.

References

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93 (2), 743–752. doi:10.3168/jds.2009-2730
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109
- Amaral, A. J., Megens, H., Crooijmans, R. P. M. A., Heuven, H. C. M., and Groenen, M. A. M. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179 (1), 569–579. doi:10.1534/genetics.107.084277
- Boichard, D., Ducrocq, V., Croiseau, P., and Fritz, S. (2016). Genomic selection in domestic animals: principles, applications and perspectives. *Comptes Rendus Biol.* 339 (7–8), 274–277. doi:10.1016/j.crv.2016.04.007
- Botelho, M. E., Lopes, M. S., Mathur, P. K., Knol, E. F., Guimarães, S. E. F., Marques, D. B. D., et al. (2021). Applying an association weight matrix in weighted genomic prediction of boar taint compounds. *J. Anim. Breed. Genet.* 138 (4), 442–453. doi:10.1111/jbg.12528
- Brøndum, R. F., Su, G., Janss, L., Sahana, G., Guldbrandsen, 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 (6), 4107–4116. doi:10.3168/jds.2014-9005
- Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol. Paris.* 42 (1), 2. doi:10.1186/1297-9686-42-2
- Cole, J. B., Dürr, J. W., and Nicolazzi, E. L. (2021). Invited review: the future of selection decisions and breeding programs: what are we breeding for, and who decides? *J. Dairy Sci.* 104 (5), 5111–5124. doi:10.3168/jds.2020-19777
- Consortium, T. B. H., Gibbs, R. A. A., Taylor, J. F. A., Van Tassell, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324 (5926), 528–532. doi:10.1126/science.1167936
- Dassonneville, R., Brndum, R. F., Druet, T., Fritz, S., Guillaume, F., Guldbrandsen, B., et al. (2011). Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* 94 (7), 3679–3686. doi:10.3168/jds.2011-4299
- de Roos, A. P. W., Hayes, B. J., and Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics* 183 (4), 1545–1553. doi:10.1534/genetics.109.104935
- Ding, X., Zhang, Z., Li, X., Wang, S., Wu, X., Sun, D., et al. (2013). Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. *J. Dairy Sci.* 96 (8), 5315–5323. doi:10.3168/jds.2012-6194
- Edel, C., Schwarzenbacher, H., Hamann, H., Neuner, S., Emmerling, R., and Götz, K. U. (2011). The German-Austrian genomic evaluation system for Fleckvieh (Simmental) cattle[J]. *Interbull Bull.* (44), 152–156.
- Farnir, F., Coppieters, W., Arranz, J., Berzi, P., Cambisano, N., Grisart, B., et al. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* 10 (2), 220–227. doi:10.1101/gr.10.2.220
- George, R., Wiggans, J. B. C. S., Hubbard, S. M., and Sonstegard, T. S. (2017). Genomic selection in dairy cattle: the USDA experience. *Annu. Rev. animal Biosci.* 5, 309–327. doi:10.1146/annurev-animal-021815-111422
- Hayes, B., Meuwissen, T., and Goddard, M. (2016). Genomic selection: a paradigm shift in animal breeding. *Anim. Front.* 6 (1), 6–14. doi:10.2527/af.2016-0002
- Hill, W. G. (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33 (2), 229–239. doi:10.1038/hdy.1974.89
- Jorjani, H., Jakobsen, J., Hjerpe, E., Palucci, V., and Dürr, J. W. (2012). Status of genomic evaluation in the Brown Swiss populations. *Interbull* (46), 46–54.
- Karaman, E., Lund, M. S., Anche, M. T., Janss, L., and Su, G. (2018). Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances

and covariances across the genome. *G3 (Bethesda, Md.)* 8 (11), 3549–3558. doi:10.1534/g3.118.200673

Legarra, A., and Reverter, A. (2019). Correction to: semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 51 (1), 69. doi:10.1186/s12711-019-0511-5

Liu, G. C. (2013). *Special history of China's dairy industry*. Beijing: China Agriculture Press.

Lund, M. S., Roos, A. P. W. D., Vries, A. G. D., Druet, T., Ducrocq, V., Fritz, S., et al. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol. Paris* 43 (1), 43. doi:10.1186/1297-9686-43-43

Lund, M. S., Su, G., Janss, L., Gulbrandsen, B., and Brøndum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. *Livest. Sci.* 166, 101–110. doi:10.1016/j.livsci.2014.05.008

Ma, Y. (2015). *Genetic variations of chinese holstein bulls in fak gene regulation zone identification and its associated with semen quality analysis* (Shandong: Shandong Normal University).

Marjanovic, J., Hulsegge, B., and Calus, M. P. L. (2021). Relatedness between numerically small Dutch Red dairy cattle populations and possibilities for multibreed genomic prediction. *J. Dairy Sci.* 104 (4), 4498–4506. doi:10.3168/jds.2020-19573

Metta, M., Kanginakudru, S., Gudiseva, N., and Nagaraju, J. (2004). Genetic characterization of the Indian cattle breeds, Ongole and Deoni (*Bos indicus*), using microsatellite markers - a preliminary study. *BMC Genet.* 5 (1), 16. doi:10.1186/1471-2156-5-16

Odani, M., Narita, A., Watanabe, T., Yokouchi, K., Sugimoto, Y., Fujita, T., et al. (2006). Genome-wide linkage disequilibrium in two Japanese beef cattle breeds. *Anim. Genet.* 37 (2), 139–144. doi:10.1111/j.1365-2052.2005.01400.x

Palombo, V., Pegolo, S., Conte, G., Cesarani, A., Macciotta, N. P. P., Stefanon, B., et al. (2021). Genomic prediction for latent variables related to milk fatty acid composition in Holstein, Simmental and Brown Swiss dairy cattle breeds. *J. Anim. Breed. Genet.* 138 (3), 389–402. doi:10.1111/jbg.12532

Park, L. (2012). Linkage disequilibrium decay and past population history in the human genome. *PLoS one* 7 (10), e46603. doi:10.1371/journal.pone.0046603

Pryce, J. E., Gredler, B., Bolormaa, S., Bowman, P. J., Egger-Danner, C., Fuerst, C., et al. (2011). Short communication: genomic selection using a multi-breed, across-country reference population. *J. Dairy Sci.* 94 (5), 2625–2630. doi:10.3168/jds.2010-3719

Steyn, Y., Lourenco, D. A. L., and Misztal, I. (2019). Genomic predictions in purebreds with a multibreed genomic relationship matrix. *J. Animal Sci.* 97 (11), 4418–4427. doi:10.1093/jas/skz296

van den Berg, I., Boichard, D., and Lund, M. S. (2016). Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. *J. Dairy Sci.* 99 (11), 8932–8945. doi:10.3168/jds.2016-11073

van den Berg, I., Meuwissen, T. H. E., Macleod, I. M., and Goddard, M. E. (2019). Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. *J. Dairy Sci.* 102 (4), 3155–3174. doi:10.3168/jds.2018-15231

Vanderick, S., Gillon, A., Glorieux, G., Mayeres, P., Mota, R., and Gengler, N. (2017). Usefulness of multi-breed models in genetic evaluation of direct and maternal calving ease in Holstein and Belgian Blue Walloon purebreds and crossbreds. *Livest. Sci.* 198 (1), 129–137. doi:10.1016/j.livsci.2017.02.019

van Grevenhof, E. M., Vandenplas, J., and Calus, M. P. L. (2019). Genomic prediction for crossbred performance using metafounders. *J. animal Sci.* 97 (2), 548–558. doi:10.1093/jas/sky433

Vanraden, P. M., Tooker, M. E., Chud, T. C. S., Norman, H. D., Megonigal, J. H., Haagen, I. W., et al. (2020). Genomic predictions for crossbred dairy cattle. *J. Dairy Sci.* 103 (2), 1620–1631. doi:10.3168/jds.2019-16634

Villar-Hernández, B. D. J., Pérez-Elizalde, S., Martini, J. W. R., Toledo, F., Perez-Rodriguez, P., Krause, M., et al. (2021). Application of multi-trait Bayesian decision theory for parental genomic selection. *G3 (Bethesda)* 11 (2), jkab012. doi:10.1093/g3journal/jkab012

Weigel, K. A. A., de Los Campos, G. A., Vazquez, A. I. A., Rosa, G. J. M., Gianola, D., and Van Tassell, C. P. (2010). Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci.* 93 (11), 5423–5435. doi:10.3168/jds.2010-3149

Weller, J. I., Ezra, E., and Ron, M. (2017). Invited review: a perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 100 (11), 8633–8644. doi:10.3168/jds.2017-12879

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18 (5), 691–699. doi:10.1093/oxfordjournals.molbev.a003851

Xu, L., Wang, Z., Zhu, B., Liu, Y., Li, H., Bordbar, F., et al. (2019). Theoretical evaluation of multi-breed genomic prediction in Chinese indigenous cattle. *Anim. (Basel)* 9 (10), 789. doi:10.3390/ani9100789

Zhang, M. H., You, Z. C., Wei, C., Li, J. F., Ge, J. J., Chen, G. L., et al. (2021). Analysis of influencing factors on production performance of Chinese Holstein cattle and Xinjiang brown cattle in Xinjiang region. *Chin. J. Animal Sci.* 57 (5), 246–252. doi:10.19556/j.0258-7033.20201113-04

Zhang, C., Dong, S., Xu, J., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, M., Luo, H., Xu, L., Shi, Y., Zhou, J., Wang, D., et al. (2022). Genomic selection for milk production traits in Xinjiang Brown cattle. *Animals* 12 (2), 136. doi:10.3390/ani12020136

Zhou, L., Heringstad, B., Su, G., Gulbrandsen, B., Meuwissen, T. H. E., Svendsen, M., et al. (2014). Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. *J. Dairy Sci.* 97 (7), 4485–4496. doi:10.3168/jds.2013-7580

Zumbach, B., Jorjani, H., and Dürr, J. (2010). Brown Swiss genomic evaluation. *Interbull Bull.* (42), 44–45.



OPEN ACCESS

EDITED BY

Xiangdong Ding,
China Agricultural University, China

REVIEWED BY

Shamik Polley,
West Bengal University of Animal and Fishery
Sciences, India
Ran Di,
Chinese Academy of Agricultural Sciences,
China

*CORRESPONDENCE

Zhihong Liu
✉ liuzh7799@163.com

RECEIVED 06 February 2024

ACCEPTED 09 April 2024

PUBLISHED 02 May 2024

CITATION

Liu Z, Qin Q, Zhang C, Xu X, Dai D, Lan M, Wang Y, Zhang J, Zhao D, Kong D, Qin T, Wu D, Gong X, Zhou X, Suhe A, Wang Z and Liu Z (2024) Effects of nonsynonymous single nucleotide polymorphisms of the *KIAA1217*, *SNTA1* and *LTBP1* genes on the growth traits of Ujumqin sheep.
Front. Vet. Sci. 11:1382897.
doi: 10.3389/fvets.2024.1382897

COPYRIGHT

© 2024 Liu, Qin, Zhang, Xu, Dai, Lan, Wang, Zhang, Zhao, Kong, Qin, Wu, Gong, Zhou, Suhe, Wang and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Effects of nonsynonymous single nucleotide polymorphisms of the *KIAA1217*, *SNTA1* and *LTBP1* genes on the growth traits of Ujumqin sheep

Zhichen Liu¹, Qing Qin¹, Chongyan Zhang¹, Xiaolong Xu¹, Dongliang Dai¹, Mingxi Lan¹, Yichuan Wang¹, Jingwen Zhang¹, Dan Zhao¹, Deqing Kong¹, Tian Qin¹, Danni Wu¹, Xuedan Gong¹, Xingyu Zhou¹, Alatan Suhe², Zhixin Wang^{1,3,4,5} and Zhihong Liu^{1,3,4,5*}

¹College of Animal Science, Inner Mongolia Agricultural University, Hohhot, China, ²East Ujumqin Banner Hersig Animal Husbandry Development Limited Liability Company, Xilin Gol League, Xilinhot, China, ³Key Laboratory of Animal Genetics, Breeding and Reproduction, Inner Mongolia Agricultural University, Hohhot, China, ⁴Key Laboratory of Mutton Sheep Genetics and Breeding, Ministry of Agriculture, Hohhot, China, ⁵Goat Genetics and Breeding Engineering Technology Research Center, Hohhot, China

Sheep body size can directly reflect the growth rates and fattening rates of sheep and is also an important index for measuring the growth performance of meat sheep. In this study, high-resolution resequencing data from four sheep breeds (Dorper sheep, Suffolk sheep, Ouessant sheep, and Shetland sheep) were analyzed. The nonsynonymous single nucleotide polymorphisms of three candidate genes (*KIAA1217*, *SNTA1*, and *LTBP1*) were also genotyped in 642 healthy Ujumqin sheep using MALDI-TOFMS and the genotyping results were associated with growth traits. The results showed that different genotypes of the *KIAA1217* g.24429511T>C locus had significant effects on the chest circumferences of Ujumqin sheep. The *SNTA1* g.62222626C>A locus had different effects on the chest depths, shoulder widths and rump widths of Ujumqin sheep. This study showed that these two sites can be used for marker-assisted selection, which will be beneficial for future precision molecular breeding.

KEYWORDS

Ujumqin sheep, nsSNP, *KIAA1217*, *SNTA1*, *LTBP1*

1 Introduction

Sheep (*Ovis aries*) are economically one of the most important animals in the world (1), providing humans with a stable and high-quality source of animal protein and animal products such as skin, hair, and milk. Since sheep were domesticated (2), they have inhabited all parts of the world following the migration of humans. In this process, both the natural environment and artificial selection have had a profound impact on the domestication of sheep, resulting in large differences in the appearances, phenotypes, and morphological structures of sheep in different regions. For example, Dorper sheep native to South Africa and Suffolk sheep (3, 4) from the United Kingdom are famous for their fast growth rates and high meat production.

The average adult sheep weight range can reach 85–120 kg. In sharp contrast, small sheep, such as Ouessant sheep and Shetland sheep (5, 6), have average adult body weights ranging from 45 to 75 kg. The growth traits of sheep (such as weight, height and length) can directly reflect the growth rates and fat growth rates of sheep and are also an important indicator of the growth performance of sheep. Therefore, understanding sheep growth traits is highly important for livestock production. At present, the common sheep varieties in China include Ujumqin sheep, Hu sheep, Xiaowei Han sheep, Tibetan sheep and others (7–10). As an excellent local variety differentiated from Mongolian sheep, Ujumqin sheep gradually developed after long-term selection and breeding and are well known for their fast growth rate and delicious meat.

In recent years, following the rapid development of biotechnology and genomics, molecular marker technology has improved. Single-nucleotide polymorphisms (SNPs) are third-generation molecular markers after restricted fragment length polymorphisms (RFLPs) and microsatellite polymorphisms (MPPs) (11). SNPs refer to DNA sequence polymorphisms caused by single nucleotide variations in the genome, including single base insertions, deletions, transitions, and transversions (12). SNPs can be classified into nonsynonymous single nucleotide polymorphisms (nsSNPs) and synonymous single nucleotide polymorphisms (sSNPs) according to their effects on gene transcription and protein translation. Nonsynonymous single nucleotide polymorphisms represent common genetic variants that alter the translated amino acid sequence, and this may affect the structure or function of the expressed protein (13). This genetic variation may lead to alterations in individual traits, such as body length and height growth traits. In this context, the results of an increasing number of studies have shown that nsSNPs have important effects on the growth traits of sheep (14, 15). By continuously mining and verifying major gene loci related to sheep economic traits, including the *MSTN* gene affecting sheep muscle growth and development, the *TBXT* gene affecting sheep tail type, and the *FGF5* candidate gene related to wool length, researchers have applied these polymorphisms of major genes in production practice through gene editing and other technologies. This approach can greatly increase the economic value of sheep; for example, biallelic knockout of *MSTN* expression in sheep was successfully achieved using CRISPR/Cas9 gene editing. Compared with wild-type (WT) sheep, *MSTN* knockout sheep have greater muscle mass and muscle fiber diameters, but their meat quality and taste are not affected (16). CRISPR/Cas9-mediated deletion of the *FGF5* gene not only results in the loss of its activity but also promotes the growth of sheep wool and cashmere, thereby increasing length and yield (17). These achievements have brought new optimism and opportunities for the development of animal husbandry. Nevertheless, the major gene loci involved in sheep growth traits still require research, especially with the deepening of our understanding of the sheep genome driven by biotechnology. There are still some potentially important gene loci that have not been revealed. Therefore, it is of great theoretical and practical significance to identify and verify the relevant sheep growth trait sites; it is expected that this investigation will provide a breakthrough for improving sheep production performance and economic benefits. It is important to identify potential genetic markers that affect the growth traits or reproductive characteristics of sheep and to accelerate the sheep breeding process.

In this study, resequencing data from four sheep breeds were divided into high-yield and low-yield groups according to body size, and population differentiation index (FST) analysis was used to identify genes and mutation sites related to the body size of the sheep. Based on the FST results, we performed GO and KEGG enrichment analyses for the top 5% of the candidate genes. Twenty-four nsSNPs across three candidate genes, *KIAA1217*, *STNA1* and *LTBP1*, were identified. However, whether these nsSNPs have an effect in Ujumqin sheep is unclear. Therefore, we collected blood from 642 Ujumqin sheep for genotyping of the above sites, matched the genotyping results with body size data, and verified the molecular markers of the main effect growth trait sites in the genomes of Ujumqin sheep.

2 Materials and methods

2.1 Sample collection

The experimental animals used in this study were bred in their country of origin and were not crossed with other sheep breeds. A total of 642 Ujumqin sheep (260 males + 382 females) were obtained from the East Ujumqin Banner breeding farm in Xilin Gol League, Inner Mongolia. All healthy animals were fed freely under the same natural conditions. Venous blood collection was performed on sheep using a vacuum blood collection vessel, and the sample storage conditions were -20°C . The body size data of the sheep were measured manually. For detailed measurement methods, see a previously published article (18).

2.2 Data sources

The raw data for this study were downloaded from the BioProject database of NCBI (PRJNA624020). The dataset contained resequencing data for four breeds of sheep (Dorper sheep, Suffolk sheep, Ouessant sheep and Shetland sheep), for a total of 37 samples.

2.3 Sequence read mapping and SNP calling

Approximately 82.86 Gb of the original sequence was obtained from each sample, with an average depth of $27.4\times$ covering the clean reads (Supplementary Datas 1, 2). Using default parameters, the 150 bp counterpart reads were mapped to the sheep reference genome Oar v.4.0 using the Burrows–Wheeler Aligner (19). The mapping results were then converted to BAM format and sorted using SAMtools (20). After mapping, we made SNP calls to all individuals using the Bayesian method implemented in SAMtools and the Genome Analysis Toolkit (GATK) (21). Then, SnpEff software (22) was used to annotate the structures of the mutation sites.

2.4 FST analysis

The four breeds of sheep were divided into two groups according to their body sizes (high-yield group: Dorper sheep, Suffolk sheep; low-yield group: Ouessant sheep, Shetland sheep). The FST value was

estimated using a 100 kb sliding window on each chromosome with a window step size of 10 kb. The average F_{ST} values are the values for the entire genomes of different populations. When the F_{ST} value of the sliding window was greater than 95% of the F_{ST} value of the genome range, it was selected as a significant window. The overlapping significant windows were then merged into segments that were considered highly differentiated regions between populations. Genes within these regions of differentiation are considered candidates for selective elimination. We used the VCFtools 0.1.16 software package (23) for data collection for the F_{ST} calculation method: (1) Calculation of in-population genetic variation: For each genetic marker, the allele frequency of each population was calculated. The degree of heterozygosity (H_o) or allelic diversity within each population was calculated. (2) Calculation of interpopulation genetic variation: The average allele frequency of all populations was calculated. These mean frequencies were used to calculate the expected heterozygosity (H_t) of the population. (3) Calculation of F_{ST} :

$$F_{ST} = \frac{(H_t - H_s)}{H_t}$$

Here, H_t is the overall expected heterozygosity between populations, and H_s is the weighted average expected heterozygosity within populations.

2.5 Enrichment analysis of key genes

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed for the top 5% of the F_{ST} genes. GO enrichment analysis was performed using the online tool g: Profiler (24), and GO classifications included biological processes (BP), cellular components (CC), and molecular functions (MF). The significantly enriched GO items were selected according to $p < 0.05$. KOBAS 3.0 (25) was used for KEGG pathway enrichment analysis, and $p < 0.05$ was considered the screening criterion for significant enrichment.

2.6 Design and synthesis of primers and classification of SNPs

Based on the F_{ST} analysis results of this study and referring to previous results (26), primer design was performed using the upstream and downstream 200 bp sequence information of nsSNPs of common genes significantly selected by XP-CLR in the top 5% of F_{ST} s. The primers were designed using the Agena online software Design Suite 2.0.¹ The designed primer sequence was derived and then synthesized by Invitrogen. The mutation sites were identified by stroma-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). The genotyping system used was the MassARRAY® MALDI-TOF system developed by Agena.

2.7 Statistical analysis

The phenotypic data of Ujumqin sheep were statistically analyzed using Excel 2021 software. SPSS 22.0 software (27) was used to analyze the correlation between genotypes and phenotypic traits. $P < 0.05$ indicated significant differences, and $p < 0.01$ indicated extremely significant differences. The following general linear model (GLM) was used for analysis: $Y = \mu + G + m + e$.

Y , trait measurement value; μ , population mean; G , genotype effect; m , sex effect; e , random residual.

2.8 nsSNP protein structure prediction

TBtools (28) was used to extract CDS information before and after mutation based on reference genome information. RNAfold² online software was used to predict the secondary structure of mutated mRNAs. SOPMA (29) was used to predict protein secondary structure. After the amino acid sequence using CDS sequence can be converted to SWISS-MODEL³ online software to forecast the mutant protein tertiary structure before and after.

3 Results

3.1 Selection of candidate genes for sheep body size

To identify candidate genes that affect sheep body size, we conducted F_{ST} selection signal analysis on the high-yield and low-yield groups of sheep (Figure 1), and a total of 1747 significant genes were selected as the top 5% of the F_{ST} . Analysis of these 1747 genes revealed that three genes (*KIAA1217*, *SNTA1* and *LTBP1*) were significantly selected by the XP-CLR method used in previous studies. Therefore, we selected nsSNPs of the above three genes for further study (Table 1).

3.2 Enrichment analysis of candidate genes

To investigate which GO enrichment terms and signaling pathways were associated with the top 5% of genes significantly enriched in the F_{ST} , we conducted GO and KEGG enrichment analyses of these 1747 genes (Figure 2). According to the GO classification statistics, 214 terms were grouped into three GO categories: cellular components, molecular functions, and biological processes. Among the three GO categories, the most significantly enriched were cytoplasm, protein binding and developmental process. GO enrichment analysis revealed that most of these genes were involved to varying degrees in the development of organisms (Figure 2A).

¹ <https://support.agenabio.com/s/online-tools>

² <http://rna.tbi.univie.ac.at/>

³ <https://swissmodel.expasy.org/>

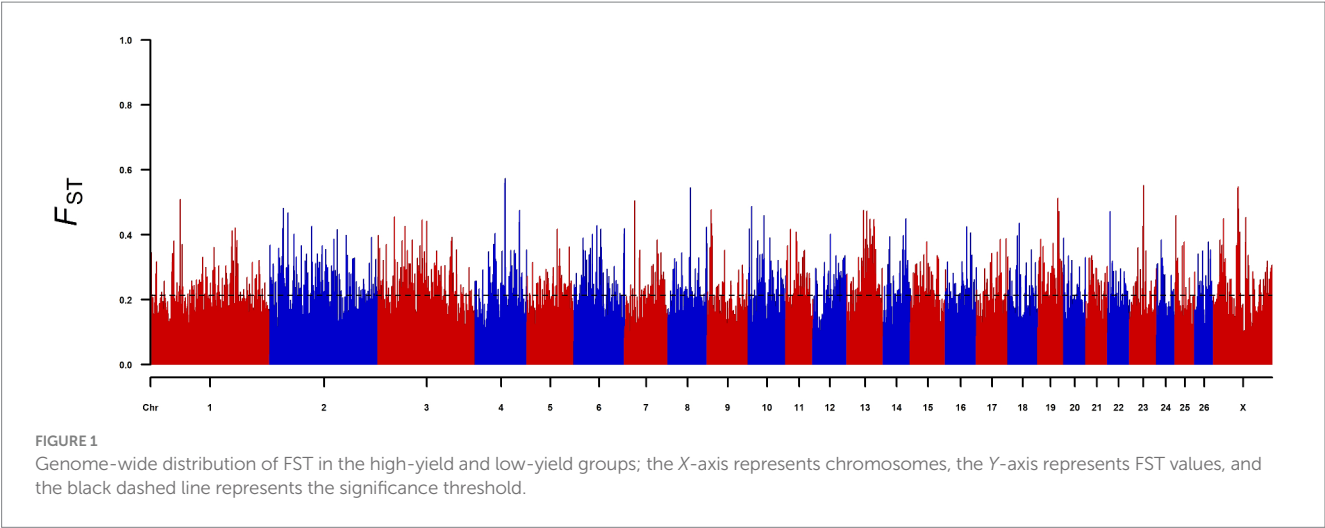


TABLE 1 The nonsynonymous single nucleotide polymorphism sites in the *KIAA1217*, *SNTA1*, and *LTBP1* genes.

SNP	Gene name	Mutation type	Position	REF	ALT
g.24086608A>C	<i>KIAA1217</i>	Missense variant	Chr13: 24086608	A	C
g.24125201C>T	<i>KIAA1217</i>	Missense variant	Chr13: 24125201	C	T
g.24125227G>A	<i>KIAA1217</i>	Missense variant	Chr13: 24125227	G	A
g.24125228T>C	<i>KIAA1217</i>	Missense variant	Chr13: 24125228	T	C
g.24125285G>A	<i>KIAA1217</i>	Missense variant	Chr13: 24125285	G	A
g.24256607C>T	<i>KIAA1217</i>	Stop gained	Chr13: 24256607	C	T
g.24256993C>T	<i>KIAA1217</i>	Missense variant	Chr13: 24256993	C	T
g.24351854A>G	<i>KIAA1217</i>	Missense variant	Chr13: 24351854	A	G
g.24387021G>A	<i>KIAA1217</i>	Missense variant	Chr13: 24387021	G	A
g.24392399A>G	<i>KIAA1217</i>	Missense variant	Chr13: 24392399	A	G
g.24399197C>A	<i>KIAA1217</i>	Missense variant	Chr13: 24399197	C	A
g.24406459C>T	<i>KIAA1217</i>	Missense variant	Chr13: 24406459	C	T
g.24424378C>T	<i>KIAA1217</i>	Missense variant	Chr13: 24424378	C	T
g.24428968T>C	<i>KIAA1217</i>	Missense variant	Chr13: 24428968	T	C
g.24429511T>C	<i>KIAA1217</i>	Missense variant	Chr13: 24429511	T	C
g.24429581T>A	<i>KIAA1217</i>	Missense variant	Chr13: 24429581	T	A
g.62220063G>A	<i>SNTA1</i>	Missense variant	Chr13: 62220063	G	A
g.62222626C>A	<i>SNTA1</i>	Missense variant	Chr13: 62222626	C	A
g.62247870T>C	<i>SNTA1</i>	Missense variant	Chr13: 62247870	T	C
g.90345529C>T	<i>LTBP1</i>	Missense variant	Chr3: 90345529	C	T
g.90377522G>A	<i>LTBP1</i>	Missense variant	Chr3: 90377522	G	A
g.90398299A>C	<i>LTBP1</i>	Missense variant	Chr3: 90398299	A	C
g.90427996G>T	<i>LTBP1</i>	Missense variant	Chr3: 90427996	G	T
g.90642790G>C	<i>LTBP1</i>	Missense variant	Chr3: 90642790	G	C

In addition, we performed KEGG pathway enrichment analysis to explore the most active pathways related to these genes. The results showed that most of the enriched pathways were related to biological development, such as the phospholipase D signaling pathway, the PI3K/Akt signaling pathway, the TGF- β signaling pathway, and cell cycle (Figure 2B).

3.3 Correlations between growth traits and nsSNPs in Ujumqin sheep

The DNA of 642 Ujumqin sheep was genotyped and the next step was to analyze the loci with more than 95% detection rate. Among the 24 nsSNPs, only 2 met the above conditions, namely, *KIAA1217*

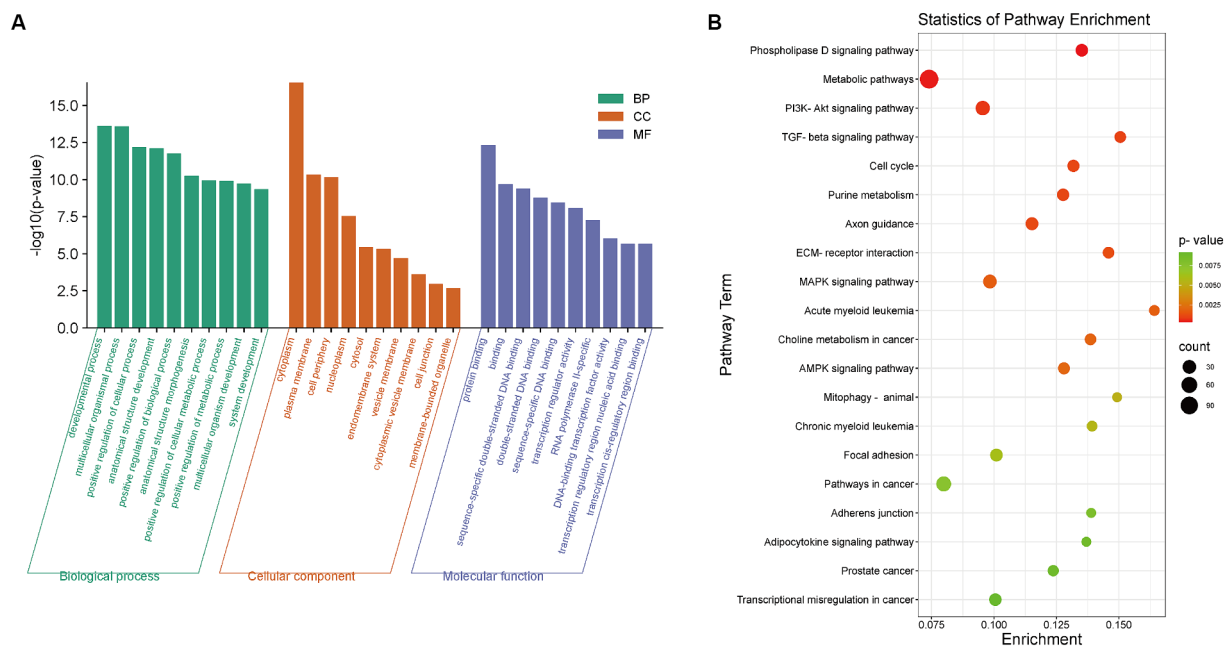


FIGURE 2

The results of GO/KEGG enrichment analysis of the top 5% of genes in the FST selection signal. (A) The results of GO enrichment analysis. Green, orange and blue represent biological processes, cell components and molecular functions in GO enrichment terms, respectively. (B) The results of KEGG enrichment analysis. The X-axis represents the ratio of differentially enriched genes to the total gene number of this term, and the Y-axis represents the enriched function/pathway. The color represents the degree of enrichment significance, and the closer to red the color is, the more significant the enrichment is. The size of the circle represents the number of genes enriched.

g.24429511T>C and *SNTA1* g.62222626C>A (Figures 3A,B). Analysis of the genotype frequencies and allele frequencies of these two loci revealed 3 genotypes at each locus. The TC genotype in the *KIAA1217* g.24429511T>C locus was the dominant genotype, and the genotype frequency was 0.48, indicating moderate polymorphism ($0.25 < \text{PIC} < 0.5$). At the *SNTA1* g.62222626C>A locus, the CC genotype was the dominant genotype, the genotype frequency was 0.84, and the polymorphism frequency was low ($\text{PIC} < 0.25$). In addition, according to the χ^2 test, both of the above two SNPs reached HWE ($p > 0.05$) (Table 2). The results of the association analysis between the typing results of the two loci and the 6-month-old body size data of Ujumqin sheep are shown in Table 3. The presence of different genotypes at the *KIAA1217* g.24429511T>C locus had significant effects on the chest circumference of Ujumqin sheep (Figure 3C), and the chest circumference of individuals with the TC genotype was significantly greater than that of individuals with the TT genotype ($p < 0.05$). The chest depths of Ujumqin sheep with an AA genotype at the *SNTA1* g.62222626C>A locus were greater than those of CC genotype and CA genotype sheep (Figure 3F), and the difference was significant ($p < 0.05$). The shoulder widths of CA genotype sheep were significantly greater than those of CC genotype sheep (Figure 3E) ($p < 0.05$), and the differences in rump width were highly significant ($p < 0.01$), as shown in Figure 3D.

3.4 Prediction of nsSNP protein structure related to growth traits of Ujumqin sheep

To better understand how the above two mutation sites affect the expression of genes involved in the growth trait changes of Ujumqin

sheep, we used TBtools to extract the CDSs of genes before and after the mutation and used online software to predict the secondary structures of mRNA and the secondary structures and tertiary structures of encoded proteins. The results showed that mutations at both sites resulted in changes in the original amino acid sequence, and mutations at *KIAA1217* g.24429511T>C resulted in a change in amino acid 1,493 from the original phenylalanine to leucine (Figures 4A,B). Amino acid 260 at *SNTA1* g.62222626C>A changed from tryptophan to leucine (Figures 4C,D). Secondary structure analysis of mRNA before and after mutation revealed that the minimum structural free energy of mRNA before and after the *KIAA1217* g.24429511T>C mutation changed from $-1983.90 \text{ kcal/mol}$ to $-1984.40 \text{ kcal/mol}$. The minimum structural free energy of the mRNA before and after the *SNTA1* g.62222626C>A site changed from -692.50 kcal/mol to -695.50 kcal/mol , and the stability increased (Supplementary Figure 1). By analyzing the changes in amino acids at the mutation sites and their effects on protein secondary structures, it was found that the proportion of α -helix and random coils increased after mutation at *KIAA1217* g.24429511T>C, and the proportion of β -turn and extended strands decreased. There was no change in the proportion of secondary structures after the *SNTA1* g.62222626C>A mutation (Table 4).

4 Discussion

Against the background of natural selection, some selective intervention behaviors of humans may cause mutations in some sites of the animal body so that the genetic traits of the organism are diverse. Phenotypic alterations are often caused by functional mutations in

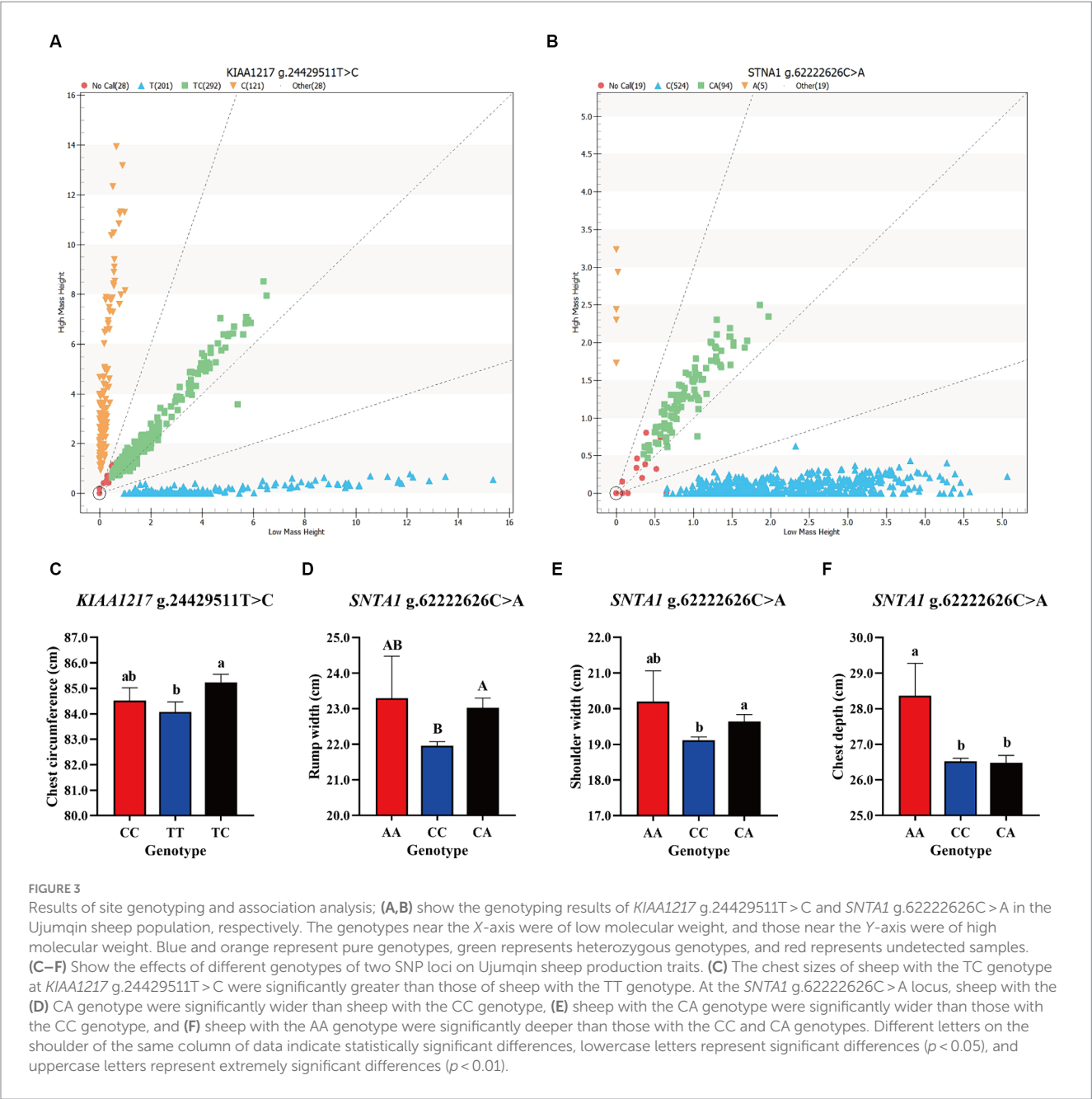


TABLE 2 Genetic parameters of nsSNPs in the experimental population of Ujumqin sheep.

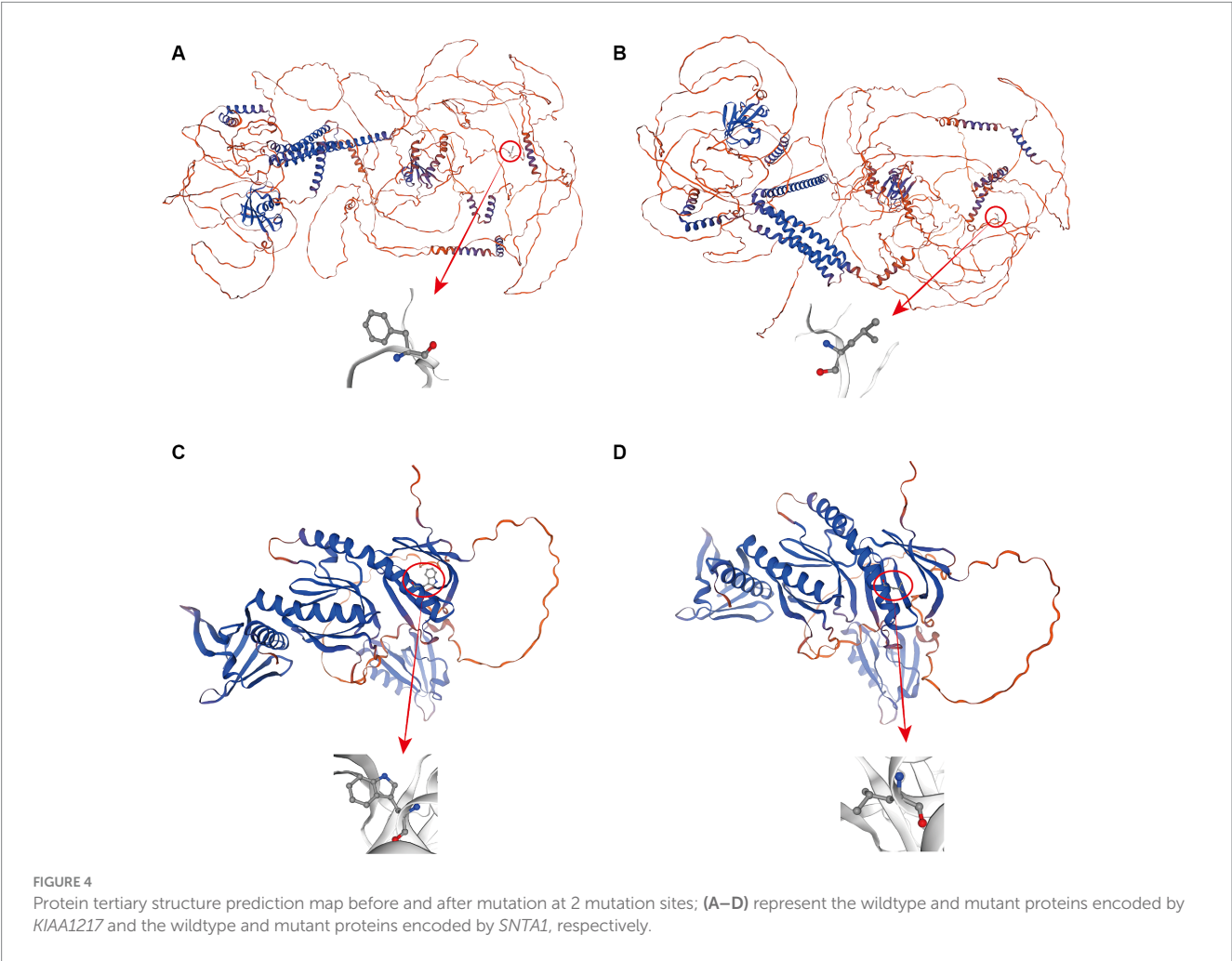
SNP	Genotype	Genotype frequency	Allele	Allele frequency	He	Ho	PIC	HWE (χ^2)
<i>KIAA1217</i> g.24429511T>C	CC (121)	0.20	T	0.57	0.48	0.52	0.35	0.62
	TT (201)	0.33	C	0.43				
	TC (292)	0.48						
<i>SNTA1</i> g.62222626C>A	AA (5)	0.01	C	0.92	0.15	0.85	0.14	0.04
	CC (524)	0.84	A	0.08				
	CA (94)	0.15						

genes that control the trait, so such mutations can serve as valid molecular markers for marker-assisted breeding (30). Among these mutations, missense mutations are considered to affect the body phenotype by affecting the mRNA expression of the related genes and the structures of the proteins. In this study, the g.24429511T>C and g.62222626C>A mutations, which are missense mutations that cause

TABLE 3 Genetic parameters of nsSNPs in the experimental population of Ujumqin sheep.

SNP	Genotype	Body slanting length	Height at wither	Hip height	Chest depth	Shoulder width	Rump width	Chest circumference
<i>KIAA1217</i> g.24429511T>C	CC (121)	63.77 ± 0.38	59.95 ± 0.36	62.18 ± 0.38	26.61 ± 0.19	18.96 ± 1.80	22.05 ± 0.24	84.52 ± 0.51 ^{ab}
	TT (201)	64.27 ± 0.29	59.86 ± 0.28	61.50 ± 0.30	26.45 ± 0.15	19.23 ± 0.14	21.97 ± 0.19	84.08 ± 0.39 ^b
	TC (292)	63.56 ± 0.24	59.87 ± 0.23	61.97 ± 0.25	26.55 ± 0.12	19.300.11	22.24 ± 0.16	85.23 ± 0.33 ^a
<i>SNTA1</i> g.62222626C>A	AA (5)	65.24 ± 1.85	60.62 ± 1.78	62.30 ± 1.86	28.37 ± 0.91 ^a	20.20 ± 0.86 ^{ab}	23.30 ± 1.18 ^{AB}	84.36 ± 2.25
	CC (524)	63.84 ± 0.18	59.85 ± 0.18	61.98 ± 0.19	26.52 ± 0.09 ^b	19.12 ± 0.09 ^b	21.96 ± 0.12 ^B	84.73 ± 0.25
	CA (94)	63.97 ± 0.43	60.26 ± 0.41	62.32 ± 0.43	26.48 ± 0.21 ^b	19.64 ± 0.20 ^a	23.03 ± 0.27 ^A	84.69 ± 0.57

Different uppercase letters in the same column represent very significant differences ($p < 0.01$), different lowercase letters represent significant differences ($p < 0.05$), and the same letters represent no significant differences.



differences in growth traits, were located in exon 19 and exon 4 of the gene, respectively. Although both sites lead to changes in amino acids encoded, there are some differences between the two mutations. The g.24429511T>C missense mutation caused the encoded amino acid to change from phenylalanine to leucine, and the amino acid properties changed from polar to nonpolar after the mutation, which may have led to changes in protein properties. Further analysis revealed that mutation of this site resulted in a decrease in the minimum structural free energy of the mRNA secondary structure, an increase in stability, an increase in the proportion of α -helix and random coils, and a decrease in the proportion of β -turn and extended strands. After the

missense mutation at g.62222626C>A, the encoded amino acid changed from tryptophan to leucine and from a polar amino acid to a nonpolar amino acid. However, from the perspective of protein secondary structure, the mutation at this site did not cause changes in the proportions of α -helix, extended strands, β -turn or random coils. The main reason may be that amino acids form extended strands in the secondary structure before and after mutation. The *SNTA1* gene encodes a cytoplasmic peripheral membrane scaffold protein. In terms of tertiary structure, changes in the amino acids of the protein may affect the structure of the protein, thus affecting the binding of related proteins and leading to the emergence of different phenotypes.

TABLE 4 Results of protein secondary structure prediction.

Gene		α -helix	β -turn	Extended strand	Random coil
<i>KIAA1217</i>	Wild type	25.21%	2.95%	9.68%	62.16%
	Mutant	25.41%	2.85%	9.52%	62.22%
<i>SNTA1</i>	Wild type	29.11%	5.74%	21.98%	43.17%
	Mutant	29.11%	5.74%	21.98%	43.17%

KIAA1217 (a sickle tail protein homolog) has a curly helix region and an actin interaction domain. The proteins it encodes are necessary for normal disc development, dendritic spine morphogenetic regulation, embryonic skeletal system development by regulating cell migration, multicellular biological development, and substrate adhesion-dependent cell proliferation. Mutations in *KIAA1217* are associated with malformations in the human backbone and tail vertebrae, and in mice, they affect the development of the spine, resulting in a reduced number of tail vertebrae and a characteristic short tail (31). Recently, researchers have shown that rapid evolution of the regulatory region of this gene in apes may lead to tail loss, which may be related to mutations in specific gene regulatory sequences (32). It has been confirmed that the vertebrae of sheep and humans are most similar in the thoracic and lumbar regions, although they show substantial differences in some dimensions (33). *SNTA1* encodes the α 1-synthetic protein, a scaffold protein that is a component of the anti-muscular dystrophin-associated protein complex (34). *SNTA1* is the link between the extracellular matrix, the intracellular signaling apparatus, and the actin cytoskeleton. *SNTA1* is involved in the regulation of the actin cytoskeleton and actin recombination (35). The *SNTA1* signaling axis plays an important role in cytoskeletal tissue (36), and researchers have found that the *SNTA1* gene is associated with amino acid and ion channel binding in different parts of bovine muscle (37). The sheep *LTBP1* gene, located on chromosome 3, encodes TGF- β -binding protein 1 and is a member of the potential TGF- β -binding protein family. One indel of the *LTBP1* gene was detected in four sheep breeds, and the effect of the *LTBP1* gene on the growth traits of small-tailed cold sheep may be related to sex (38). Cao et al. conducted genome-wide DNA methylation sequencing in a subpopulation of Chinese Mongolian sheep and found that DNA methylation in three regions and two CpG sites in *LTBP1* was significantly correlated with its RNA expression, and this gene was also identified as a potential candidate gene associated with weight variation (39). Unfortunately, 5 nsSNPs of the *LTBP1* gene were not detected in the population genotyping results of Ujumqin sheep in this study. Moreover, exploring the influence of the *LTBP1* gene on the growth traits of Ujumqin sheep will also constitute a direction of follow-up research.

In recent years, following the development of biotechnology and genome research, an increasing number of SNPs related to the growth traits of livestock and poultry have been discovered. Studies have shown that the g.3148C>T polymorphism of the *SIRT1* gene affects the heart circumference of Tibetan sheep, and the g.8074T>A polymorphism of the *SIRT2* gene is significantly correlated with body weight and body length (40). Cao et al. discovered OARX_76354330.1 and s64890.1 to be functional SNPs for growth traits of Hu sheep through genome-wide association analysis of body weight and

identified three candidate genes related to body weight in Hu sheep (41): *CAPN6*, *ITGA11* and *SCMH1*. By using genome-wide high-density SNP data (600 K) for selective scanning tests for important phenotypic traits, researchers have identified genes related to sheep body size, such as *RMI1* and *SCD5* (42). In the selection of sheep genotypes, it is necessary to comprehensively select SNP genotypes corresponding to the target traits; we therefore need to use different methods to identify additional SNP sites that affect sheep growth traits. Overall, this study used resequencing data from different sheep breeds to mine candidate genes and nsSNPs affecting sheep body size and verified their roles in Ujumqin sheep, thereby obtaining two molecular markers that could be applied in production practice. However, it is not yet clear whether these sites are equally useful in other sheep breeds, and this will be an important direction for our future research. Our results provide new insights into the mining of SNPs related to sheep growth and provide new genetic markers for the genetic improvement in Ujumqin sheep.

5 Conclusion

In this study, resequencing data from four sheep breeds were used to identify three genes related to the body size of sheep. Twenty-four nsSNPs across three genes were identified in Ujumqin sheep. Association analysis revealed that different genotypes of two SNP loci had differences in chest circumference, chest depth, body width and caudal breadth, and mRNA secondary structure stability was enhanced after mutation. These SNPs can be used as a molecular marker for Ujumqin sheep breeding and lay a foundation for future precise molecular breeding.

Data availability statement

The data used in this study is sourced from the BioProject database of NCBI, numbered PRJNA624020.

Ethics statement

The animal study was approved by in this study, blood was collected in accordance with the International Guiding Principles for Biomedical Research involving animals and approved by the Special Committee on Scientific Research and Academic Ethics of Inner Mongolia Agricultural University, responsible for the approval of Biomedical Research Ethics of Inner Mongolia Agricultural University [Approval No. (2020) 056]. No specific permissions were required for these activities, and no endangered or protected species were involved. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

ZCL: Writing – original draft, Writing – review & editing. QQ: Writing – original draft, Writing – review & editing. CZ: Writing – original draft, Writing – review & editing. XX: Writing – original draft, Writing – review & editing. DD: Writing – original draft, Writing – review & editing. ML: Writing – original draft, Writing

– review & editing. YW: Writing – original draft, Writing – review & editing. JZ: Writing – original draft, Writing – review & editing. DZ: Writing – original draft, Writing – review & editing. DK: Writing – original draft, Writing – review & editing. TQ: Writing – original draft, Writing – review & editing. DW: Writing – original draft, Writing – review & editing. XG: Writing – original draft, Writing – review & editing. XZ: Writing – original draft, Writing – review & editing. AS: Writing – original draft, Writing – review & editing. ZW: Writing – original draft, Writing – review & editing. ZHL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Project of Northern Agriculture and Livestock Husbandry Technical Innovation Center, Chinese Academy of Agricultural Sciences (BFGJ2022002); The National Key R&D Program of China (2021YFD1200901); National Natural Science Foundation of China (32060742); The Major Science and Technology Projects of Inner Mongolia Autonomous Region (2020ZD0004); The central government guides local science and technology development funds (2022ZY0185); Higher educational Basic scientific research business fee project of Inner Mongolia Autonomous Region (BR221005); and University Youth Science and Technology Talent Support Project of Inner Mongolia Autonomous Region (NJYT23012).

References

- Hiendleder S, Kaupé B, Wassmuth R, Janke A. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc Biol Sci.* (2002) 269:893–904. doi: 10.1098/rspb.2002.1975
- Zeder MA. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *Proc Natl Acad Sci USA.* (2008) 105:11597–604. doi: 10.1073/pnas.0801317105
- Qiao G, Xu P, Guo T, Wu Y, Lu X, Zhang Q, et al. Genetic basis of Dorper sheep (*Ovis Aries*) revealed by Long-read De novo genome assembly. *Front Genet.* (2022) 13:846449. doi: 10.3389/fgene.2022.846449
- Fitzmaurice S, Conington J, Fetherstone N, Pabiou T, McDermott K, Wall E, et al. Genetic analyses of live weight and carcass composition traits in purebred Texel, Suffolk and Charollais lambs. *Animal.* (2020) 14:899–909. doi: 10.1017/S1751731119002908
- Smit JJ, Schonewille JT, Beynen AC. Transient lowering of plasma vitamin B12 concentrations in Ouessant sheep fed on a potassium-rich ration. *Int J Vitamin Nutr Res Int Zeitschrift für Vitamin.* (1999) 69:273–6. doi: 10.1024/0300-9831.69.4.273
- Thomas DL. Performance and utilization of northern European short-tailed breeds of sheep and their crosses in North America: a review. *Animal.* (2010) 4:1283–96. doi: 10.1017/S1751731110000856
- Li S, Luo R, Lai D, Ma M, Hao F, Qi X, et al. Whole-genome resequencing of Ujumqin sheep to investigate the determinants of the multi-vertebral trait. *Genome.* (2018) 61:653–61. doi: 10.1139/gen-2017-0267
- Li Y, Chen Z, Fang Y, Cao C, Zhang Z, Pan Y, et al. Runs of homozygosity revealed reproductive traits of Hu sheep. *Genes.* (2022) 13:1848. doi: 10.3390/genes13101848
- Luo M, Wang L, Xiao C, Zhou M, Li M, Li H. Mir136 regulates proliferation and differentiation of small tail Han sheep preadipocytes. *Adipocytes.* (2023) 12:2173966. doi: 10.1080/21623945.2023.2173966
- Jing X, Zhou J, Degen A, Wang W, Guo Y, Kang J, et al. Comparison between Tibetan and small-tailed Han sheep in adipocyte phenotype, lipid metabolism and energy homeostasis regulation of adipose tissues when consuming diets of different energy levels. *Br J Nutr.* (2020) 124:668–80. doi: 10.1017/S0007114520001701
- Gao L, Jia J, Kong X. A Snp-based molecular barcode for characterization of common wheat. *PLoS One.* (2016) 11:e0150947. doi: 10.1371/journal.pone.0150947
- Sun Y, Lan X, Lei C, Zhang C, Chen H. Haplotype combination of the bovine Cfd2 gene sequence variants and association with growth traits in Qinchuan cattle. *Gene.* (2015) 563:136–41. doi: 10.1016/j.gene.2015.03.016
- Nakken S, Alseth I, Rognes T. Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes. *Neuroscience.* (2007) 145:1273–9. doi: 10.1016/j.neuroscience.2006.09.004
- Zhang L, Ma X, Xuan J, Wang H, Yuan Z, Wu M, et al. Identification of Mef2b and Trhd gene polymorphisms related to growth traits in a new Ujumqin sheep population. *PLoS One.* (2016) 11:e0159504. doi: 10.1371/journal.pone.0159504
- Li X, Ding N, Zhang Z, Tian D, Han B, Liu S, et al. Identification of somatostatin receptor subtype 1 (Sstr1) gene polymorphism and their association with growth traits in Hulun Buir sheep. *Genes.* (2021) 13:77. doi: 10.3390/genes13010077
- Zhou S, Kalds P, Luo Q, Sun K, Zhao X, Gao Y, et al. Optimized Cas9:sgRNA delivery efficiently generates biallelic MSTN knockout sheep without affecting meat quality. *BMC Genomics.* (2022) 23:348. doi: 10.1186/s12864-022-08594-6
- Li WR, Liu CX, Zhang XM, Chen L, Peng XR, He SG, et al. Crispr/Cas9-mediated loss of Fgf5 function increases wool staple length in sheep. *FEBS J.* (2017) 284:2764–73. doi: 10.1111/febs.14144
- Qin Q, Dai D, Zhang C, Zhao C, Liu Z, Xu X, et al. Identification of body size characteristic points based on the mask R-Cnn and correlation with body weight in Ujumqin sheep. *Front Vet Sci.* (2022) 9:995724. doi: 10.3389/fvets.2022.995724
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and Samtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a Mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010) 20:1297–303. doi: 10.1101/gr.107524.110
- Cingolani P. Variant annotation and functional prediction: Snpeff. *Methods Mol Biol.* (2022) 2493:289–314. doi: 10.1007/978-1-0716-2293-3_19
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and Vcftools. *Bioinformatics.* (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* (2019) 47:W191–8. doi: 10.1093/nar/gkz369

Acknowledgments

The author thanks the staff of Ujumqin sheep breeding farm in East Ujumqin Banner, Xilin Gol League, Inner Mongolia for their help.

Conflict of interest

AS was employed by East Ujumqin Banner Hersig animal husbandry development limited liability 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.2024.1382897/full#supplementary-material>

25. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, et al. Kobas-I: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* (2021) 49:W317–25. doi: 10.1093/nar/gkab447
26. Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun.* (2020) 11:2815. doi: 10.1038/s41467-020-16485-1
27. Duricki DA, Soleman S, Moon LD. Analysis of longitudinal data from animals with missing values using Spss. *Nat Protoc.* (2016) 11:1112–29. doi: 10.1038/nprot.2016.048
28. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. Tbttools-ii: a "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol Plant.* (2023) 16:1733–42. doi: 10.1016/j.molp.2023.09.010
29. Geourjon C, Deléage G. Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* (1995) 11:681–4. doi: 10.1093/bioinformatics/11.6.681
30. Dekkers JC, Hospital F. The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet.* (2002) 3:22–32. doi: 10.1038/nrg701
31. Al Dhaheer N, Wu N, Zhao S, Wu Z, Blank RD, Zhang J, et al. Kiaa1217: a novel candidate gene associated with isolated and syndromic vertebral malformations. *Am J Med Genet A.* (2020) 182:1664–72. doi: 10.1002/ajmg.a.61607
32. Shao Y, Zhou L, Li F, Zhao L, Zhang BL, Shao F, et al. Phylogenomic analyses provide insights into primate evolution. *Science.* (2023) 380:913–24. doi: 10.1126/science.abn6919
33. Wilke HJ, Kettler A, Wenger KH, Claes LE. Anatomy of the sheep spine and its comparison to the human spine. *Anat Rec.* (1997) 247:542–55. doi: 10.1002/(SICI)1097-0185(199704)247:4<542::AID-AR13>3.0.CO;2-P
34. Dong T, Zhang S, Chang Y, Bai R, Jiang Y, Ma S, et al. The establishment of a homozygous Snta1 knockout human embryonic stem cell line (Wae009-a-50) using the Crispr/Cas9 system. *Stem Cell Res.* (2021) 51:102196. doi: 10.1016/j.scr.2021.102196
35. Bhat SS, Parray AA, Mushtaq U, Fazili KM, Khanday FA. Actin Depolymerization mediated loss of Snta1 phosphorylation and Rac1 activity has implications on Ros production, cell migration and apoptosis. *Apoptosis.* (2016) 21:737–48. doi: 10.1007/s10495-016-1241-6
36. Ali R, Mir HA, Hamid R, Bhat B, Shah RA, Khanday FA, et al. Actin modulation regulates the Alpha-1-Syntrophin/P66shc mediated redox signaling contributing to the RhoA Gtpase protein activation in breast cancer cells. *Front Oncol.* (2022) 12:841303. doi: 10.3389/fonc.2022.841303
37. Zhang T, Wang T, Niu Q, Zheng X, Li H, Gao X, et al. Comparative transcriptomic analysis reveals region-specific expression patterns in different beef cuts. *BMC Genomics.* (2022) 23:387. doi: 10.1186/s12864-022-08527-3
38. Wang S, Yi X, Wu M, Zhao H, Liu S, Pan Y, et al. Detection of key gene Indels in Tgf-B pathway and its relationship with growth traits in four sheep breeds. *Anim Biotechnol.* (2021) 32:194–204. doi: 10.1080/10495398.2019.1675682
39. Cao J, Wei C, Liu D, Wang H, Wu M, Xie Z, et al. DNA methylation landscape of body size variation in sheep. *Sci Rep.* (2015) 5:13950. doi: 10.1038/srep13950
40. Gui LS, Raza SHA, Zhou L, Garcia M, Abd El-Aziz AH, Wei D, et al. Association between single nucleotide polymorphisms in Sirt1 and Sirt2 loci and growth in Tibetan sheep. *Animals.* (2020) 10:1362. doi: 10.3390/ani10081362
41. Cao Y, Song X, Shan H, Jiang J, Xiong P, Wu J, et al. Genome-wide association study of body weights in Hu sheep and population verification of related single-nucleotide polymorphisms. *Front Genet.* (2020) 11:588. doi: 10.3389/fgene.2020.00588
42. Xu SS, Gao L, Shen M, Lyu F. Whole-genome selective scans detect genes associated with important phenotypic traits in sheep (*Ovis Aries*). *Front Genet.* (2021) 12:738879. doi: 10.3389/fgene.2021.738879



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Rugang Tian,
Inner Mongolia Academy of Agricultural and
Animal Husbandry Sciences, China
Zhixin Chai,
Southwest Minzu University, China

*CORRESPONDENCE

Ming Zhang,
✉ zhangming@sicau.edu.cn

RECEIVED 05 February 2024

ACCEPTED 17 April 2024

PUBLISHED 30 May 2024

CITATION

Zhang S, Li J, Zhao Y, Tang Y, Li H, Song T, An T,
Guan J, Li X and Zhang M (2024), Whole-
genome resequencing reveals genetic diversity,
differentiation, and selection signatures of yak
breeds/populations in southwestern China.
Front. Genet. 15:1382128.
doi: 10.3389/fgene.2024.1382128

COPYRIGHT

© 2024 Zhang, Li, Zhao, Tang, Li, Song, An,
Guan, Li and Zhang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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.

Whole-genome resequencing reveals genetic diversity, differentiation, and selection signatures of yak breeds/populations in southwestern China

Shilin Zhang¹, Jing Li¹, Yanhua Zhao¹, Yujun Tang¹, Hao Li¹,
Tianzeng Song², Tianwu An³, Jiuqiang Guan³, Xiaowei Li⁴ and
Ming Zhang^{1,5,6*}

¹College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China, ²Institute of Animal Science, Tibet Academy of Agricultural and Animal Husbandry Science, Lhasa, China, ³Sichuan Academy of Grassland Science, Chengdu, China, ⁴Breeding Farm of Longri, Agriculture and Rural Bureau of Aba Prefecture in Sichuan, Hongyuan, China, ⁵Key Laboratory of Livestock and Poultry Multi-Omics, Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China, ⁶Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, China

The Sichuan-Yunnan region is the main production area of yaks in southwestern China, with rich genetic resources of Yaks. Nevertheless, there have been limited study on the genetic characteristics of the entire yak populations in Tibet and southwestern China. In this study, we performed whole-genome resequencing to identify genetic variation information in a total of 198 individuals from six yak breeds (populations) in Sichuan (Muli yak, Jinchuan yak, Changtai yak, Maiwa yak), Yunnan (Zhongdian yak), and Tibet (Tibetan yak). The aim was to investigate the whole-genome genetic diversity, population genetic structure, and genome selection signatures. We observed that all six populations exhibit abundant genetic diversity. Except for Tibetan yaks, which showed low nucleotide diversity (0.00104), the remaining yak populations generally displayed high nucleotide diversity (0.00129–0.00153). Population genetic structure analysis revealed that, among the six yak populations, Muli yak exhibited greater differentiation from other yak populations and formed a distinct cluster independently. The Maiwa yak population displayed a complex genetic structure and exhibited gene exchange with Jinchuan and Changtai yaks. Positive selection signals were detected in candidate genes associated with growth (*GNB4*, *HMG2*, *TRPS1*, and *LTBP1*), reproduction (*PI4KB*, *DYNC111*, and *GRIP1*), immunity (*CD200* and *IL1RAP*), lactation (*SNX13* and *CPM*), hypoxia adaptation (*NDUFB6*, *PRKN*, and *MRPS9*), hair (*KRT24*, *KRT25*, and *KRT26*), meat quality (*SUCLG2*), digestion and absorption (*CLDN1*), and pigment deposition (*OCA2*) using the integrated P_i and F_{ST} methods. This study provides significant insights into understanding the whole-genome genetic characteristics of yak populations in Tibet and southwestern China.

KEYWORDS

Bos grunniens, whole-genome resequencing, genomic diversity, population genetic structure, selection signature

Introduction

Yaks are a unique livestock species in the plateau region, displaying high-altitude adaptation to extreme and harsh environments, including cold and hypoxia (Wiener et al., 2003; Aguiar et al., 2018; Qi et al., 2019). They hold significant importance in local animal husbandry, providing meat, milk, wool, labor, fuel, and other essential resources to the local herders (Wiener et al., 2003). Additionally, yaks serve as a valuable gene bank for genetic studies (Qiu et al., 2012).

The exploration of population genetic structure of yak has become a prominent field in domestic animal. Studies have been conducted on the genetic diversity of yaks using microsatellites and mitochondrial DNA (Qiu et al., 2015; Wang et al., 2021; Hameed et al., 2022; Ma et al., 2022). Moreover, the assembly and publication of the first yak reference genome in 2012, followed by subsequent improvements, have resulted in a highly complete and accurate yak reference genome at the chromosome level (Qiu et al., 2012; Ji et al., 2021; Zhang et al., 2021). This achievement has laid a robust foundation for investigating genetic variation at the whole-genome level in yaks, leading to the widespread use of whole-genome resequencing technology. This technology has enabled analyses of genetic variation in yaks, covering aspects such as their origin, domestication (Medugorac et al., 2017; Chai et al., 2020; Meng et al., 2022), genetic diversity and structure (Wang et al., 2014; Zhang et al., 2016; Lan et al., 2018; Wang et al., 2019; Li et al., 2022), high-altitude adaptation (Qiu et al., 2012; Qiu et al., 2015; Guang Xin et al., 2019b; Lan et al., 2021; Gao et al., 2022) and selection pressures (Lan et al., 2018; Xie et al., 2018; Guang Xin et al., 2019a; Li et al., 2022).

Southwestern China, specifically Sichuan Province's Garze Tibetan Autonomous Prefecture, Aba Tibetan and Qiang Autonomous Prefecture, and Yunnan Province's Diqing Tibetan Autonomous Prefecture, serves as the primary production area for yaks (Lan et al., 2018). In this region, several yak breeds, including Jiulong yak, Maiwa yak, Muli yak, Jinchuan yak, Changtai yak, and Zhongdian yak, have been officially recognized as excellent indigenous breeds by the government (Lan et al., 2021). Each of these yak breeds possesses unique characteristics and outstanding traits. For example, the Jinchuan yak stands out with approximately 52% of individuals having 15 thoracic vertebrae and 15 pairs of ribs, one more thoracic vertebra and one more pair of ribs than other yak breeds (Lan et al., 2018). Additionally, multi-ribbed Jinchuan yaks demonstrate higher meat and milk production, as well as superior reproductive performance compared to other breeds (Mipam et al., 2012). Maiwa yaks, on the other hand, produce milk with a protein content 40%–60% higher than that of native bovine milk (Chen et al., 2021). Protecting, developing, and utilizing the genetic resources of yak populations in this region holds great significance for the advancement of China's yak industry. However, comprehensive studies on genetic variation detection, and selection signal analysis at the whole-genome level for multiple yak breeds in this region remain unreported.

To address the genetic gap, this study detected genetic variation, elucidate the genetic diversity, explore population genetic structure and genome-wide selection using six yak populations including four from Sichuan (Maiwa yak, Muli yak, Jinchuan yak, and Changtai yak), one from Yunnan (Zhongdian yak), and a collection of Tibetan

yak populations. Our findings provide a scientific and theoretical basis for the protection, development, and genetic improvement of yak genetic resources in Tibet and southwestern China.

Materials and methods

Sample collection and whole-genome resequencing

In this study, we collected a total of 69 blood samples from Maiwa yak (MW) (The altitude of each sampling location ranges from 3400 m to 3600 m), 20 from Zhongdian yak (ZD), and 59 from Tibetan yak (XZ). Additionally, we collected 18 tissue samples each from Muli yak (ML) and Changtai yak (CT), and 14 tissue samples from Jinchuan yak (JC). The Tibetan yak population includes 16 Niangya yaks (NY), 11 Chawula yaks (CWL), 6 Sibuyaks (SB), 17 Dangxiong yaks (DX), 3 Naqu yaks (NQ), and 6 Pali yaks (PL) (Supplementary Figure S1; Supplementary Table S1). Genomic DNA was extracted using the TIANamp Genomic DNA Kit (Tiangen Biotech Co., Ltd., Beijing, China). 59 qualified Tibetan yak genomic DNA samples were sent to Biomarker Technologies (Beijing, China) for paired-end sequencing on the Illumina HiSeq 2500 platform. The remaining yak genomic DNA samples were sent to BGI Genomics (Shenzhen, China) for paired-end sequencing on the T7 platform, with each read length of 150 bp.

Read mapping and SNP calling

The raw data were filtered through the quality control process conducted by fastp 0.20.0 (Chen et al., 2018). The high-quality clean reads were mapped to the latest yak reference genome (BosGru3.0, GCA_005887515.2) using Burrows–Wheeler Aligner (BWA, v0.7.8-r455) software and SAMtools (Li and Durbin, 2009) to get the original mapping results stored in BAM format. Then, the results were dislodged duplication by Picard (<http://broadinstitute.github.io/picard/>). After genome mapping we undertook SNP calling for all individuals using SamTools, v1.13, and The “mpileup” command was executed to identify SNPs with the parameters as -m 2 -F 0.002 -d 1000 -u -C 50 (Li et al., 2015). The genotype data were filtered using PLINK 1.9 based on minor allele frequencies, missing genotype rates, and biallelic alleles with parameter: -geno 0.2 --maf 0.05 --biallelic-only. SnpEff (Cingolani et al., 2012) was used for functional annotation of variants. The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) (Li et al., 2021) in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation (CNCB-NGDC Members and Partners, 2022), under accession number GVM000675 (<https://bigd.big.ac.cn/gvm/getProjectDetail?Project=GVM000675>).

Population genetic diversity analysis

We estimated the genomic nucleotide diversity (π) of each yak population using VCFtools (v0.1.16) (Danecek et al., 2011) with the

parameters (-window-pi 50,000 -window-pi-step 20,000). VCFtools was also used to calculate the average observed heterozygosity (H_O) and expected heterozygosity (H_E) as well as the average inbreeding coefficient (F) for six yak populations using the -het parameter. The pattern of linkage disequilibrium (LD) among 198 yaks from six populations can be effectively reflected by calculating the coefficient of determination (r^2) between pairwise SNPs using PopLDdecay (Zhang et al., 2019) with default parameter.

Population genetic structure and phylogenetic tree analysis

Linkage sites in the genomic data were removed with parameters (-indep-pair-wise 50 10 0.2) using PLINK (v1.9) software. The genetic structure was analyzed using ADMIXTURE (v1.3.0) and the optimal clustering was also calculated to determine the best K value using ADMIXTURE’s cross-validation procedure. The smartpca module in EIGENSOFT (Patterson et al., 2006) was used to perform principal component analysis (PCA) on six yaks populations. The results were plotted using the ggplot2 package in R for the first principal component (PC1) and the second principal component (PC2). Based on the pairwise distance matrix among samples, a neighbor-joining (NJ) tree was constructed by MEGA (v11.0.11). The original tree file was visualized and beautified using the ggtree package in R. VCFtools was used to calculate the fixation index (F_{ST}) between six yak populations.

Genome-wide selective sweep and functional enrichment analysis

Based on the results of F_{ST} , we used the Muli yak as the control group and the Maiwa yak as the selection group to perform selection sweep based on two analysis methods, nucleotide diversity (π), and the genome-wide distribution of pairwise fixation index (F_{ST}). π and F_{ST} was calculated with 50 kb sliding windows and 20 kb steps between adjacent windows using the VCFtools. The windows with the top 5% F_{ST} and -log10 π values were considered as candidate region under strong selection and the candidate genes in the regions were identified. The intersection of two analysis results was taken as the final selected genes and a Venn diagram was drawn using the online website jvenn (<http://jvenn.toulouse.inra.fr/app/index.html>). The functions of the selected genes were further explored by conducting the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/anno_iden.php).

Results

Genome resequencing and genetic variation

Sequencing of 59 Tibetan yaks yielded 967.591 Gb of raw data and 953.379 Gb of filtered clean data. An additional 139 yaks were sequenced, producing 5274.237 Gb of raw data and 5084.898 Gb of filtered clean data (Supplementary Table S2). After mapping to the yak reference genome (BosGru3.0), the sequencing depth of the

TABLE 1 SNP functional annotation results.

Category		Number of SNPs	Proportion (%)
Upstream		29816	0.453
UTR3		33663	0.512
UTR5		7299	0.111
Exonic	Stop gain	270	0.004
	Stop loss	58	0.001
	Synonymous	15846	0.241
	Non-synonymous	14609	0.222
	Unknown	9211	0.140
Intronic		2226657	33.843
Splicing		334	0.005
Downstream		36667	0.557
Upstream/downstream		1045	0.016
Intergenic		4203881	63.895

59 individual Tibetan yaks ranged from 3.45 to 4.31×, while the sequencing depth of the remaining 139 yaks ranged from 9.12 to 25.42× (Supplementary Table S3). The mapping rate for all samples was above 96.68%. In total, we obtained 44,262,798 high-quality SNPs. The highest number of SNPs (5,145,209) was detected in Tibetan yaks, followed by Changtai yaks (3,007,140), Jinchuan yaks (2,955,926), Muli yaks (2,951,598), and Maiwa yaks (2,764,492), while the lowest number of SNPs was detected in Zhongdian yaks (2,640,551). Most SNPs were distributed in intergenic and intronic regions, with 63.90% located in intergenic regions and 33.84% in intronic regions. Only 0.61% of SNPs were distributed in exonic regions. Of the SNPs in exonic regions, 39.62% were synonymous and 36.53% were non-synonymous (Table 1).

Population genomic diversity

π among the six breeds of yaks ranged from 0.00104 to 0.00153 (Figure 1A). Changtai yak exhibited the highest π at 0.00153, while Tibetan yak had the lowest at 0.00104. Apart from Tibetan yak, which showed relatively low π , the other yak populations generally exhibited high π , ranging from 0.00129 to 0.00153, (Supplementary Table S4). Except for Tibetan yak (0.05707), the F of the other five yak breeds was relatively low (−0.07969 to 0.01778) (Supplementary Table S4, Figure 1B). Muli yak had the lowest average inbreeding coefficient at −0.07969. The LD analysis showed that Muli yak had the highest average LD (r^2), followed by Jinchuan yak, Changtai yak, Zhongdian yak, and Maiwa yak. In contrast, Tibetan yaks had the lowest average LD (r^2) (Figure 1C). Results for heterozygosity indicated that the differences between observed heterozygosity (H_O) and expected heterozygosity (H_E) among the six breeds of yaks were small (Figure 1D). Muli yak had the highest H_O and H_E , at 0.36546 and 0.33848, respectively, while Tibetan yak had the lowest H_O and H_E , at 0.27363 and 0.29020, respectively (Supplementary Table S4). This suggests that Tibetan yak had the fastest LD decay

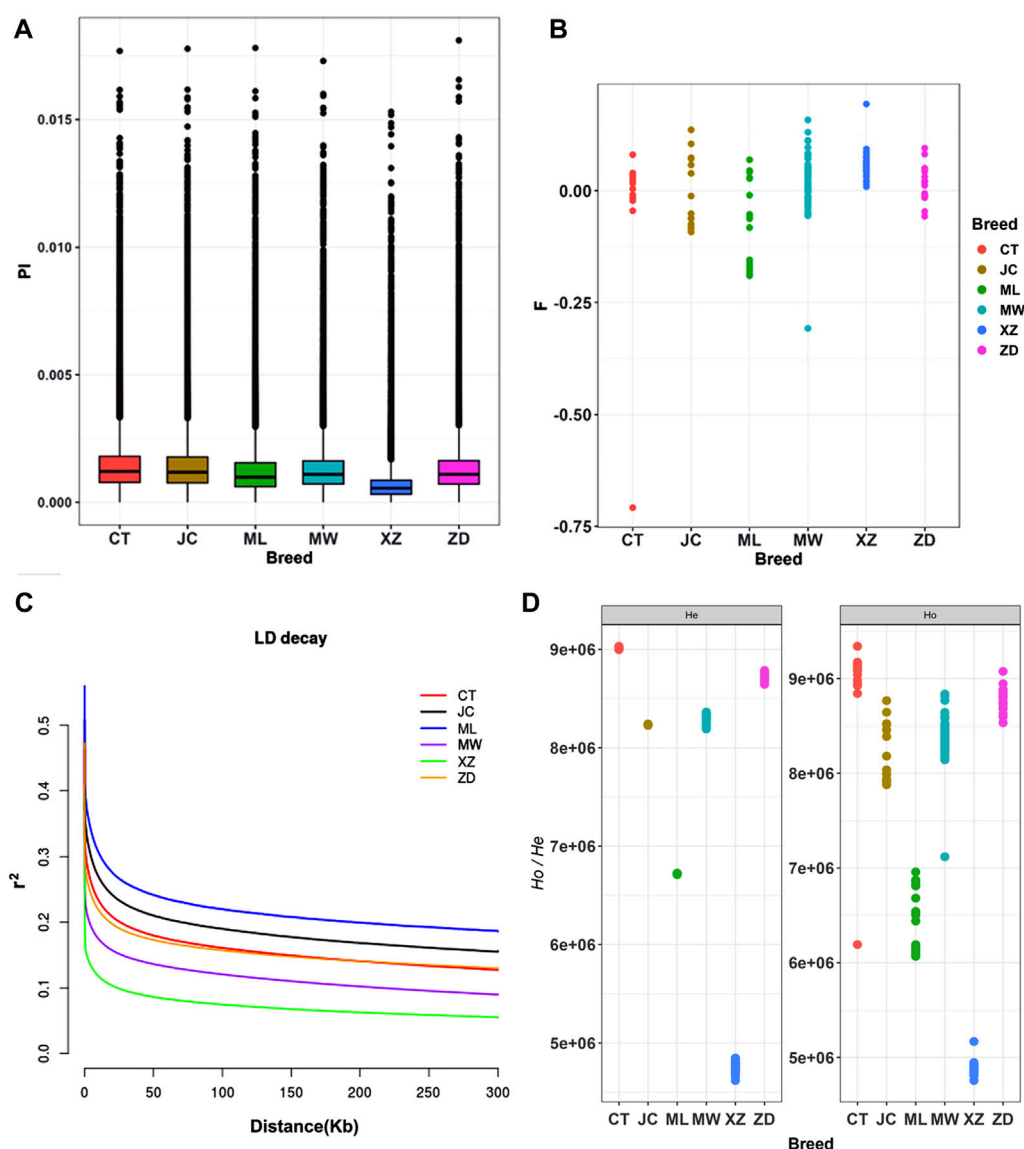


FIGURE 1 Analysis of genomic diversity, P_i , F and LD for these yak breeds/populations in southwestern China. **(A)** Box plots of nucleotide diversity (P_i). **(B)** the average inbreeding coefficient (F). **(C)** Decay of linkage disequilibrium (LD). **(D)** The observed heterozygosity (H_o) and expected heterozygosity (H_e): CT: Changtai yak, JC: Jinchuan yak, ML: Muli yak, MW: Maiwa yak, XZ: Tibet yak, ZD: Zhongdian yak.

rate among these six breeds, while Muli yak had the slowest decay rate.

Population genetic structure

To examine the population genetic structure and relationships among six different yak breeds, we conducted a series of genome-wide analyses, including population genetic structure, PCA, and phylogenetic analysis. The results of the ancestral component analysis showed that the K value was equal to 3 and the lowest value of the cross-validation (CV) error was reached, proving that the optimal population grouping was when the ancestor population was 3 (Figure 2A). Moreover, when $K = 2$, the six yak breeds were grouped into two ancestral populations, distinguishing Muli yak

from the other yak breeds. When $K = 3$, a new ancestral population emerged, further distinguishing Muli yak, Tibetan yak, and the other yak breeds. Most Muli yak samples have a relatively homogeneous ancestral population, while Tibetan yak showed a small number of mixed third ancestry. The majority of the other yak breeds showed a composition of three ancestries. When $K = 4$, the overall ancestry composition of each breed remained consistent with the results at $K = 3$ (Figure 2B). However, there was an uneven distribution phenomenon in the ancestry composition ratio of each sample of Maiwa yak.

According to the PCA results (Figure 2C), we observed that the Muli yak population formed a distinct and separate cluster from the other yak populations. On the other hand, most Maiwa yaks overlapped with the Jinchuan yaks, Changtai yaks, Tibetan yaks, and Zhongdian yaks, forming a clustered group together.

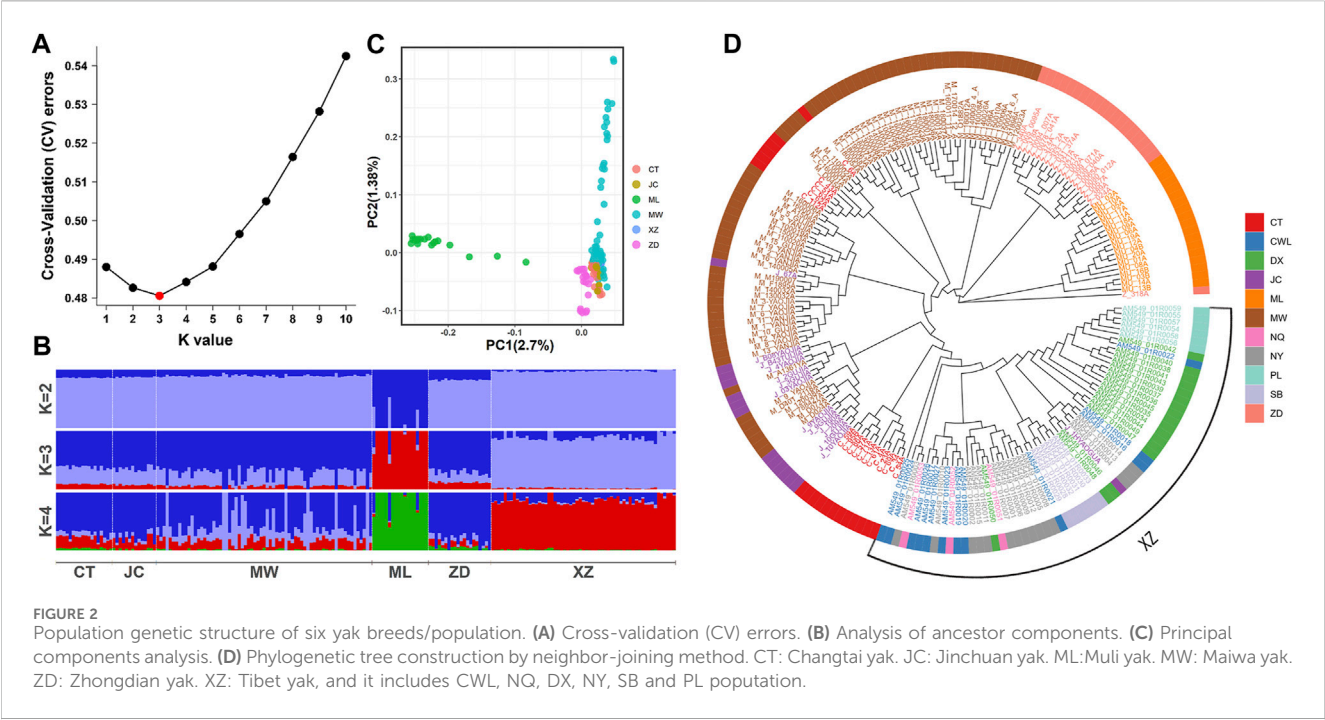


TABLE 2 F_{ST} values between pairwise populations.

Breeds/Population	JC	ML	MW	XZ	ZD	Mean
Changtai (CT)	0.0342	0.0868	0.0204	0.0315	0.0317	0.0409
Jinchuan (JC)		0.0999	0.0286	0.0422	0.0444	0.0499
Muli (ML)			0.0777	0.0998	0.0848	0.0898
Maiwa (MW)				0.0216	0.0241	0.0345
Tibet (XZ)					0.0371	0.0464
Zhongdian (ZD)						0.0444

Based on genome-wide data, the NJ tree revealed that Muli yaks and Zhongdian yaks were grouped into two distinct branches, with Maiwa yaks forming a separate branch (Figure 2D). Some individuals from the Changtai yak and Jinchuan yak populations clustered together within the Maiwa yak branch, indicating close genetic relatedness. Similarly, some individuals from the Changtai yak and Tibetan yak populations formed another branch within the Maiwa yak group, suggesting a close genetic relationship between them. Notably, one individual from the Zhongdian yak population exhibited significant hybridization and formed a separated branch from the main population to form its own branch. Furthermore, one individual from the Jinchuan yak population clustered with the Tibetan yak population. Within the Tibetan yak population, the Pali yaks displayed a relatively pure bloodline, reflecting their distant geographic distribution from other Tibetan yaks.

The average F_{ST} of Muli yak was found to be the highest at 0.0898 (Table 2), indicating a moderate degree of genetic differentiation between Muli yak and other yak populations. On

the other hand, the average fixation index of other yak populations was relatively low, with Maiwa yak having the lowest average differentiation index at 0.0345.

Genome-wide selective sweeps and functional enrichment analysis of candidate genes

According to the results of genetic differentiation between populations, we observed that the average F_{ST} between Muli and Maiwa yak was the highest, indicating a moderate level of differentiation. Consequently, we designated Muli yak as the control group and Maiwa yak as the selection group for the selection sweep analysis using two methods, P_i and F_{ST} . Subsequently, we identified the top 5% of scanning results as candidate regions (Figures 3A, B). As a result, a total of 1935 genes (P_i) and 2344 genes (F_{ST}) were identified, with 187 genes overlapping between the two analyses (Figure 3C). These candidate genes were subject to strong selection and

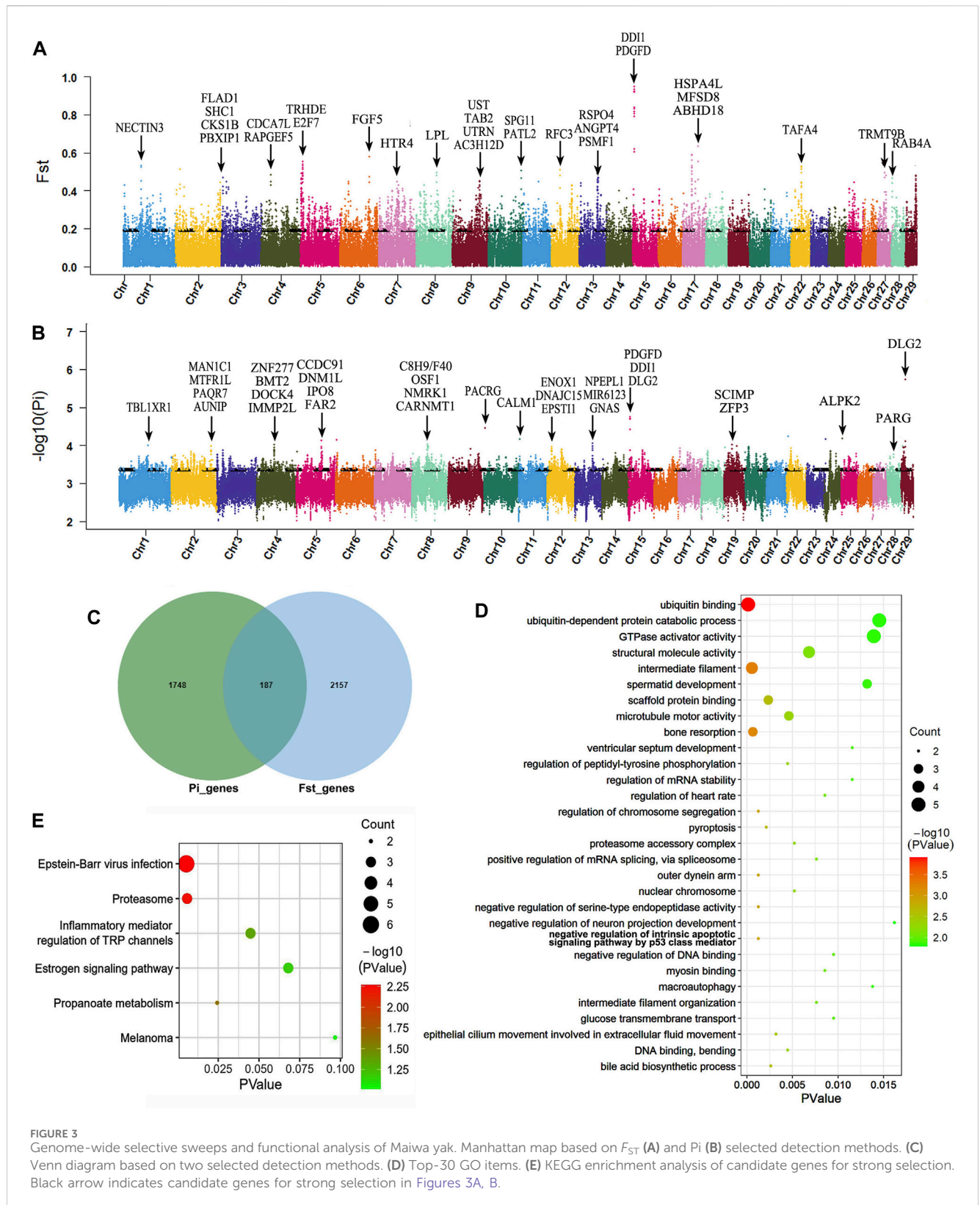


FIGURE 3

Genome-wide selective sweeps and functional analysis of Maiwa yak. Manhattan map based on F_{ST} (A) and Π (B) selected detection methods. (C) Venn diagram based on two selected detection methods. (D) Top-30 GO items. (E) KEGG enrichment analysis of candidate genes for strong selection. Black arrow indicates candidate genes for strong selection in Figures 3A, B.

mainly participated in biological processes such as growth (*GNB4*, *HMG2*, *TRPS1*, and *LTBP1*), reproduction (*PI4KB*, *DYNCH11*, and *GRIP1*), immunity (*CD200* and *IL1RAP*), neural development (*ST3GAL3*, *KIAA0319L*, and *SFPQ*), lipid metabolism and accumulation (*CYP27A1*, *AGMO*, and

KBTD2), lactation (*SNX13* and *CPM*), bone development (*ZNF687* and *PPP2R2B*), hypoxia adaptation (*NDUFB6*, *PRKN*, and *MRPS9*), hair development (*KRT24*, *KRT25*, and *KRT26*), meat quality (*SUCLG2*), digestion and absorption (*CLDN1*), and pigment deposition (*OCA2*).

Functional enrichment analysis using Gene Ontology (GO) was performed on candidate genes, revealing GO terms with a significance level of $p < 0.05$ (Figure 3D). The analysis showed that the 187 selected genes were predominantly enriched in functions associated with immune regulation, reproductive regulation, regulation of heart rate, cell apoptosis, mitosis, digestive synthesis and metabolism, transmembrane substance transport, epithelial cell proliferation and wound healing, hormone secretion and regulation, bone resorption, heme transport, and ventricular septum development. Furthermore, KEGG pathway enrichment analysis was performed, identifying pathways with a significance level of $p < 0.1$ (Figure 3E). The identified pathways were linked to virus infection, propanoate metabolism, proteasome, inflammatory mediator regulation of TRP channels, estrogen signaling, and melanoma.

Discussion

Genetic diversity reflects the adaptability of a population to its environment and its evolutionary history. It holds significant value for the protection and further utilization of animal genetic resources. In this study, we observed that the nucleotide diversity of Tibetan yak is low, likely due to their isolated geographical location, leading to limited genetic interchange with other populations. Chai et al. (2020) classified 91 domestic yaks from 31 populations into three groups based on their phylogenetic relationships. They noted that the nucleotide diversity and genetic diversity of the yak population in the central Tibetan region were both low, aligning with our study results.

The observed heterozygosity (Ho) of Muli yak, Jinchuan yak, and Changtai yak was higher than the expected heterozygosity, and the inbreeding coefficient was generally lower. This suggests that these populations abundant exhibit genetic diversity and a low degree of inbreeding. This may be attributed to the large-scale purebred breeding and hybrid improvement efforts carried out on yaks in Sichuan in recent years. As a result of this work, several yak varieties have been developed for breeding populations and core breeding groups. On the other hand, the observed heterozygosity of Tibetan yak was low, while the inbreeding coefficient was high. This suggests a lack of genetic diversity, possibly due to the traditional feeding practices in the region that have led to significant inbreeding and a decline in reproductive performance. The LD decay analysis results indicated that Muli yak, Jinchuan yak, and Changtai yak have experienced more intense artificial selection in recent years, as evidenced by their slower LD decay rate, as evidenced by their slower LD decay rate. In contrast, Tibetan yak exhibited a faster LD decay compared to other populations, indicating less human intervention. This aligns with previous LD decay analysis results for Jinchuan yak and Tibetan yak (Wang et al., 2020).

In previous studies, Lan et al. (2018) investigated the population genetic structure of Jinchuan yaks at the whole-genome level. The results from phylogenetic tree and PCA analysis both indicated that Jinchuan yaks form an independent branch within the domestic yak population and are distantly related to Zhongdian yak, Tibetan yak, and Qinghai yak. Additionally, the ancestral component analysis results revealed that when $K = 3$, the Jinchuan yak population was separated from the rest of the domestic yak population. In the present study, the ancestral component analysis, PCA analysis, and

F_{ST} statistics results of the six yak populations all pointed towards Muli yak exhibiting a relatively independent population genetic structure and distant relationship with other yak populations. Most of the Maiwa yak showed a high degree of hybridization, characterized by the smallest average differentiation index and a complex genetic background. This genetic variation is likely influenced by the diverse geographical environment of the two primary yak breeding regions. Muli yak and Maiwa yak are classified as the “Hengduan Mountains type” and “Qinghai-Tibet Plateau type,” respectively, based on their ecological characteristics. Muli yak predominantly inhabits mountain valleys, leading to grassland fragmentation and restricted gene flow, resulting in isolated breeding populations with artificial selection. Conversely, Maiwa yak is primarily found in high-altitude terrains with expansive grasslands, where primitive year-round grazing is common, leading to a lower degree of artificial selection. It is speculated that the Maiwa yak population in this study may have originated from two distinct large populations due to the establishment of a core breeding group for Maiwa yak and the implementation of various effective breeding practices. One population may have introduced Jinchuan and Changtai yak bloodlines, while the other population retained a relatively pure genetic lineage without external influences.

The identification of selective signatures in a population is crucial for understanding its evolutionary history and economic traits. A total of 187 candidate positively selected genes were identified in the Maiwa yak population, and their potential effects on relevant important traits were discussed. For example, the *CLDN1* gene, a component of endothelial tight junctions, has been demonstrated to decrease intestinal permeability and improve immune adaptability in calves through the upregulation of its expression (Ghaffari et al., 2021). It also plays an effective role in repairing the rumen epithelial barrier of slow-growing yaks (Hu et al., 2019). Thus, the *CLDN1* gene may be associated with the stronger gastrointestinal immune function of Maiwa yaks, enabling them to adapt to the challenging high-altitude wetland grassland environment. The *GNB4* gene can control cattle growth by influencing the expression of growth-related hormones in the pituitary gland (Lu et al., 2020). The *HMG2* gene, a transcription factor of insulin-like growth factor gene 2 (IGF2), plays a significant role in the growth trait of navel length in beef cattle (Aguilar et al., 2018). Similarly, the *TRPS1* gene is significantly associated with the increase in both cattle body weight and eye muscle area (Lee et al., 2011). Therefore, the *GNB4*, *HMG2*, and *TRPS1* genes are likely to be closely associated with the growth and development of Maiwa yak.

This study also identified several genes that are specifically associated with reproduction. For instance, *PI4KB* is responsible for regulating actin aggregation during sperm capacitation (Etkovitz et al., 2007). *DYNC111* participates in transporting molecules and organelles during oocyte maturation and is the most prevalent transport system in cells (Racedo et al., 2008). *GRIPI* has been validated as a marker for detecting cow estrus (Hlaing et al., 2001; Lee et al., 2017). *CARPT* is a novel regulatory factor for ovarian function during follicular wave emergence in cows and is thought to play a potential role in dominant follicle selection during the follicular wave period in single-ovulating species as an “ovulation quota gatekeeper” (Mihm et al., 2000; Lv et al., 2009; Smith et al., 2010). These genes may have an impact on maintaining a stable reproductive pattern of Maiwa yak in the harsh high-altitude environments.

The challenging high-altitude environment and extensive production practices have resulted in the strong selection of genes associated with immunity and adaptation to hypoxia in the genome of Maiwa yak. For example, the reduction of the *CD200* gene has a significant effect on the quantity of pulmonary epithelial cells in radiation conditions (Beach et al., 2023). This finding might elucidate the relevant mechanism that enables the Maiwa yak to adapt to high-altitude with intense radiation and maintain lung function without harm. The *IL1RAP* gene has been identified as a central gene encoding the interleukin-1 receptor accessory protein, which plays a significant role in the immune system (Dos Santos Silva et al., 2019). Furthermore, three genes, *NDUFB6*, *PRKN*, and *MRPS9*, are closely linked to mitochondrial structure and normal function (Buggiotti et al., 2021; Fang et al., 2022; Zhang et al., 2022). Their influence on respiratory and energy metabolism efficiency may be associated with Maiwa yak's adaptation to high-altitude hypoxic environments.

In addition, this study identified lactation-related genes *SNX13* and *CPM* among the selected genes of Maiwa yak. The *SNX13* gene is highly correlated with milk production traits in cows, while the *CPM* gene has a significant effect on the fatty acid content in cow milk (Rincón et al., 2009; Shi et al., 2020). The *SUCLG2* gene has been identified as a candidate biomarker for beef tenderness in cattle (Li and Li, 2021).

Three keratin family genes, *KRT24*, *KRT25*, and *KRT26*, have been identified. The *KRT24* gene plays a crucial role in hair growth and follicle development in yaks (Bao et al., 2022). The long and dense wool of Maiwa yak helps them retain heat and withstand cold weather at high altitudes. It is worth noting that a pigment deposition-related gene, *OCA2*, related to pigment deposition has been identified, potentially influencing black pigment deposition in Maiwa yak's wool (Grønskov et al., 2007; Caduff et al., 2017; Ballan et al., 2022).

Conclusion

In conclusion, this study systematically investigates the genetic diversity, population genetic structure, kinship relationships, genetic differentiation, and gene selection of several yak breeds (populations) in Tibet and southwestern China at the whole-genome level. Overall, all six populations exhibit rich genetic diversity. Muli yak shows higher differentiation from other yak populations and forms an independent cluster. The Maiwa yak population demonstrates a complex genetic structure. The candidate genes identified for Maiwa yak are associated with various functions such as growth, reproduction, immunity, lactation, hypoxia adaptation, hair development, meat quality, digestion and absorption, and pigment deposition. These findings provide a scientific and theoretical foundation for the exploration and conservation of yak genetic resources in this region, as well as for the selection and breeding of yak breeds.

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://bigd.big.ac.cn/gym/getProjectDetail?Project=GVM000675>, GVM000675.

Ethics statement

The animal studies were approved by the Animal Welfare and Ethics Committee of Sichuan Agricultural University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

SZ: Data curation, Formal Analysis, Methodology, Writing—original draft, Software. JL: Data curation, Methodology, Investigation, Writing—original draft. YZ: Writing—original draft, Data curation, Resources. YT: Data curation, Methodology, Writing—review and editing. HL: Data curation, Methodology, Writing—review and editing. TS: Data curation, Methodology, Resources, Writing—review and editing. TA: Data curation, Resources, Writing—review and editing. JG: Data curation, Resources, Writing—review and editing. XL: Data curation, Resources, Writing—review and editing. MZ: Conceptualization, Funding acquisition, Supervision, Writing—review and editing, Project administration.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by National Key Research and Development Projects of China (2021YFD1600201).

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.2024.1382128/full#supplementary-material>

References

- Aguiar, T. S., Torrecilha, R. B. P., Milanesi, M., Utsunomiya, A. T. H., Trigo, B. B., Tijjani, A., et al. (2018). Association of copy number variation at intron 3 of HMG2 with navel length in *Bos indicus*. *Front. Genet.* 9, 627. doi:10.3389/fgene.2018.00627
- Ballan, M., Bovo, S., Schiavo, G., Schiavitto, M., Negrini, R., and Fontanesi, L. (2022). Genomic diversity and signatures of selection in meat and fancy rabbit breeds based on high-density marker data. *Genet. Sel. Evol.* 54 (1), 3. doi:10.1186/s12711-022-00696-9
- Bao, Q., Zhang, X., Bao, P., Liang, C., Guo, X., Yin, M., et al. (2022). Genome-wide identification, characterization, and expression analysis of keratin genes (KRTs) family in yak (*Bos grunniens*). *Gene* 818, 146247. doi:10.1016/j.gene.2022.146247
- Beach, T. A., Finkelstein, J. N., and Chang, P. Y. (2023). Epithelial responses in radiation-induced lung injury (RILI) allow chronic inflammation and fibrogenesis. *Radiat. Res.* doi:10.1667/rade-22-00130.1
- Buggiotti, L., Cheng, Z., Salavati, M., Wathes, C. D., and Genotype plus Environment Consortium (2021). Comparison of the transcriptome in circulating leukocytes in early lactation between primiparous and multiparous cows provides evidence for age-related changes. *BMC Genomics* 22 (1), 693. doi:10.1186/s12864-021-07977-5
- Caduff, M., Bauer, A., Jagannathan, V., and Leeb, T. (2017). OCA2 splice site variant in German Spitz dogs with oculocutaneous albinism. *PLoS One* 12 (10), e0185944. doi:10.1371/journal.pone.0185944
- Chai, Z. X., Xin, J. W., Zhang, C. F., Dawayangla, Luosang, Zhang, Q., et al. (2020). Whole-genome resequencing provides insights into the evolution and divergence of the native domestic yaks of the Qinghai-Tibet Plateau. *BMC Evol. Biol.* 20 (1), 137. doi:10.1186/s12862-020-01702-8
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560
- Chen, Y., Qu, S., Huang, Z., Ren, Y., Wang, L., and Rankin, S. A. (2021). Analysis and comparison of key proteins in Maiwa yak and bovine milk using high-performance liquid chromatography mass spectrometry. *J. Dairy Sci.* 104 (8), 8661–8672. doi:10.3168/jds.2021-20269
- Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly. (Austin)* 6 (2), 80–92. doi:10.4161/fly.19695
- CNCB-NGDC Members and Partners (2022). Database resources of the national genomics data center, China national center for bioinformation in 2022. *Nucleic Acids Res.* 50, D27–D38. doi:10.1093/nar/gkab951
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330
- Dos Santos Silva, D. B., Fonseca, L. F. S., Pinheiro, D. G., Muniz, M. M. M., Magalhães, A. F. B., Baldi, F., et al. (2019). Prediction of hub genes associated with intramuscular fat content in Nelore cattle. *BMC Genomics* 20 (1), 520. doi:10.1186/s12864-019-5904-x
- Etkovitz, N., Rubinstein, S., Daniel, L., and Breitbart, H. (2007). Role of PI3-kinase and PI4-kinase in actin polymerization during bovine sperm capacitation. *Biol. Reprod.* 77 (2), 263–273. doi:10.1095/biolreprod.106.056705
- Fang, Z., Liu, G., Zhu, M., Wang, S., Jiang, Q., Looor, J. J., et al. (2022). Low abundance of mitophagy markers is associated with reactive oxygen species overproduction in cows with fatty liver and causes reactive oxygen species overproduction and lipid accumulation in calf hepatocytes. *J. Dairy Sci.* 105 (9), 7829–7841. doi:10.3168/jds.2021-21774
- Gao, X., Wang, S., Wang, Y. F., Li, S., Wu, S. X., Yan, R. G., et al. (2022). Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak. *Nat. Commun.* 13 (1), 4887. doi:10.1038/s41467-022-32164-9
- Ghaffari, M. H., Sadri, H., Steinhoff-Wagner, J., Hammon, H. M., and Sauerwein, H. (2021). Effects of colostrum feeding on the mRNA abundance of genes related to toll-like receptors, key antimicrobial defense molecules, and tight junctions in the small intestine of neonatal dairy calves. *J. Dairy Sci.* 104 (9), 10363–10373. doi:10.3168/jds.2021-20386
- Gronskov, K., Ek, J., and Brondum-Nielsen, K. (2007). Oculocutaneous albinism. *Orphanet J. Rare Dis.* 2, 43. doi:10.1186/1750-1172-2-43
- Guang Xin, E., Basang, W. D., and Zhu, Y. B. (2019b). Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. *J. Anim. Breed. Genet.* 136 (5), 371–377. doi:10.1111/jbg.12403
- Guang Xin, E., Yang, B. G., Basang, W. D., Zhu, Y. B., An, T. W., and Luo, X. L. (2019a). Screening for signatures of selection of Tianzhu white yak using genome-wide re-sequencing. *Anim. Genet.* 50 (5), 534–538. doi:10.1111/age.12817
- Hameed, A., Schlecht, E., Tariq, M., Buerkert, A., Scheper, C., König, S., et al. (2022). Phenotypic and genetic diversity of domestic yak (*Bos grunniens*) in high-altitude rangelands of Gilgit-Baltistan, Pakistan. *J. Anim. Breed. Genet.* 139 (6), 723–737. doi:10.1111/jbg.12730
- Hlaing, M., Nam, K., Lou, J., Pope, W. F., and Nephew, K. P. (2001). Evidence for expression of estrogen receptor cofactor messenger ribonucleic acid in the ovary and uterus of domesticated animals (sheep, cow and pig). *Life Sci.* 68 (12), 1427–1438. doi:10.1016/s0024-3205(01)00937-7
- Hu, R., Zou, H., Wang, Z., Cao, B., Peng, Q., Jing, X., et al. (2019). Nutritional interventions improved rumen functions and promoted compensatory growth of growth-retarded yaks as revealed by integrated transcripts and microbiome analyses. *Front. Microbiol.* 10, 318. doi:10.3389/fmicb.2019.00318
- Ji, Q. M., Xin, J. W., Chai, Z. X., Zhang, C. F., Dawa, Y., Luo, S., et al. (2021). A chromosome-scale reference genome and genome-wide genetic variations elucidate adaptation in yak. *Mol. Ecol. Resour.* 21 (1), 201–211. doi:10.1111/1755-0998.13236
- Lan, D., Ji, W., Xiong, X., Liang, Q., Yao, W., Mipam, T. D., et al. (2021). Population genome of the newly discovered jinchuan yak to understand its adaptive evolution in extreme environments and generation mechanism of the multirib trait. *Integr. Zool.* 16 (5), 685–695. doi:10.1111/1749-4877.12484
- Lan, D., Xiong, X., Mipam, T. D., Fu, C., Li, Q., Ai, Y., et al. (2018). Genetic diversity, molecular phylogeny, and selection evidence of jinchuan yak revealed by whole-genome resequencing. *G3 (Bethesda)* 8 (3), 945–952. doi:10.1534/g3.118.300572
- Lee, S. H., van der Werf, J. H., Kim, N. K., Lee, S. H., Gondro, C., Park, E. W., et al. (2011). QTL and gene expression analyses identify genes affecting carcass weight and marbling on BTA14 in Hanwoo (Korean Cattle). *Mamm. Genome* 22 (9–10), 589–601. doi:10.1007/s00335-011-9331-9
- Lee, W. Y., Park, M. H., Kim, K. W., Song, H., Kim, K. B., Lee, C. S., et al. (2017). Identification of lactoferrin and glutamate receptor-interacting protein 1 in bovine cervical mucus: a putative marker for oestrous detection. *Reprod. Domest. Anim.* 52 (1), 16–23. doi:10.1111/rda.12744
- Li, C., Tian, D., Tang, B., Liu, X., Teng, X., Zhao, W., et al. (2021). Genome Variation Map: a worldwide collection of genome variations across multiple species. *Nucleic Acids Res.* 49, D1186–D1191. doi:10.1093/nar/gkaa1005
- Li, G., Luo, J., Wang, F., Xu, D., Ahmed, Z., Chen, S., et al. (2022). Whole-genome resequencing reveals genetic diversity, differentiation, and selection signatures of yak breeds/populations in Qinghai, China. *Front. Genet.* 13, 1034094. doi:10.3389/fgene.2022.1034094
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, K. X., Hong, W., Jiao, H. W., Wang, G. D., Rodriguez, K. A., Buffenstein, R., et al. (2015). Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *PNAS* 112 (38), 11905–11910. doi:10.1073/pnas.1514896112
- Li, S., and Li, C. (2021). Proteomics discovery of protein biomarkers linked to yak meat tenderness as determined by label-free mass spectrometry. *Anim. Sci. J.* 92 (1), e13669. doi:10.1111/asj.13669
- Lu, X., Arbab, A. A. I., Zhang, Z., Fan, Y., Han, Z., Gao, Q., et al. (2020). Comparative transcriptomic analysis of the pituitary gland between cattle breeds differing in growth: yunling cattle and leiqiong cattle. *Anim. (Basel)* 10 (8), 1271. doi:10.3390/ani10081271
- Lv, L., Jimenez-Krassel, F., Sen, A., Bettgowda, A., Mondal, M., Folger, J. K., et al. (2009). Evidence supporting a role for cocaine- and amphetamine-regulated transcript (CARTPT) in control of granulosa cell estradiol production associated with dominant follicle selection in cattle. *Biol. Reprod.* 81 (3), 580–586. doi:10.1095/biolreprod.109.077586
- Ma, Z. J., Li, G. Z., Chen, S. M., Han, J. L., and Hanif, Q. (2022). Rich maternal and paternal genetic diversity and divergent lineage composition in wild yak (*Bos mutus*). *Anim. Biotechnol.* 33 (6), 1382–1386. doi:10.1080/10495398.2021.1884567
- Medugorac, I., Graf, A., Grohs, C., Rothhammer, S., Zagdsuren, Y., Gladys, E., et al. (2017). Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* 49 (3), 470–475. doi:10.1038/ng.3775
- Meng, G., Bao, Q., Ma, X., Chu, M., Huang, C., Guo, X., et al. (2022). Analysis of copy number variation in the whole genome of normal-haired and long-haired tianzhu white yaks. *Genes (Basel)* 13 (12), 2405. doi:10.3390/genes13122405
- Mihm, M., Austin, E. J., Good, T. E., Ireland, J. L., Knight, P. G., Roche, J. F., et al. (2000). Identification of potential intrafollicular factors involved in selection of dominant follicles in heifers. *Biol. Reprod.* 63 (3), 811–819. doi:10.1095/biolreprod.63.3.811
- Mipam, T. D., Wen, Y., Fu, C., Li, S., Zhao, H., Ai, Y., et al. (2012). Maternal phylogeny of a newly-found yak population in China. *Int. J. Mol. Sci.* 13 (9), 11455–11470. doi:10.3390/ijms130911455
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2 (12), e190. doi:10.1371/journal.pgen.0020190
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178. doi:10.1101/201178

- Qi, X., Zhang, Q., He, Y., Yang, L., Zhang, X., Shi, P., et al. (2019). The transcriptomic landscape of yaks reveals molecular pathways for high altitude adaptation. *Genome Biol. Evol.* 11 (1), 72–85. doi:10.1093/gbe/evy264
- Qiu, Q., Wang, L., Wang, K., Yang, Y., Ma, T., Wang, Z., et al. (2015). Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat. Commun.* 6, 10283. doi:10.1038/ncomms10283
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The yak genome and adaptation to life at high altitude. *Nat. Genet.* 44 (8), 946–949. doi:10.1038/ng.2343
- Racedo, S. E., Wrenzycki, C., Herrmann, D., Salamone, D., and Niemann, H. (2008). Effects of follicle size and stages of maturation on mRNA expression in bovine *in vitro* matured oocytes. *Mol. Reprod. Dev.* 75 (1), 17–25. doi:10.1002/mrd.20770
- Rincón, G., Islas-Trejo, A., Casellas, J., Ronin, Y., Soller, M., Lipkin, E., et al. (2009). Fine mapping and association analysis of a quantitative trait locus for milk production traits on *Bos taurus* autosome 4. *J. Dairy Sci.* 92 (2), 758–764. doi:10.3168/jds.2008-1395
- Shi, L., Liu, L., Lv, X., Ma, Z., Li, C., Li, Y., et al. (2020). Identification of genetic effects and potential causal polymorphisms of CPM gene impacting milk fatty acid traits in Chinese Holstein. *Anim. Genet.* 51 (4), 491–501. doi:10.1111/age.12936
- Smith, G. W., Sen, A., Folger, J. K., and Ireland, J. J. (2010). Putative role of cocaine- and amphetamine-regulated transcript (CARTPT) in dominant follicle selection in cattle. *Soc. Reprod. Fertil. Suppl.* 67, 105–117. doi:10.5661/rdr-vii-105
- Wang, H., Chai, Z., Hu, D., Ji, Q., Xin, J., Zhang, C., et al. (2019). A global analysis of CNVs in diverse yak populations using whole-genome resequencing. *BMC Genomics* 20 (1), 61. doi:10.1186/s12864-019-5451-5
- Wang, K., Hu, Q., Ma, H., Wang, L., Yang, Y., Luo, W., et al. (2014). Genome-wide variation within and between wild and domestic yak. *Mol. Ecol. Resour.* 14 (4), 794–801. doi:10.1111/1755-0998.12226
- Wang, X., Pei, J., Bao, P., Cao, M., Guo, S., Song, R., et al. (2021). Mitogenomic diversity and phylogeny analysis of yak (*Bos grunniens*). *BMC Genomics* 22 (1), 325. doi:10.1186/s12864-021-07650-x
- Wang, Y., Cai, H., Luo, X., Ai, Y., Jiang, M., and Wen, Y. (2020). Insight into unique somitogenesis of yak (*Bos grunniens*) with one additional thoracic vertebra. *BMC Genomics* 21 (1), 201. doi:10.1186/s12864-020-6598-9
- Wiener, G., Han, J., and Long, R. (2003). *The yak FAO regional office for asia and the pacific*.
- Xie, X., Yang, Y., Ren, Q., Ding, X., Bao, P., Yan, B., et al. (2018). Accumulation of deleterious mutations in the domestic yak genome. *Anim. Genet.* 49 (5), 384–392. doi:10.1111/age.12703
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875
- Zhang, S., Liu, W., Liu, X., Du, X., Zhang, K., Zhang, Y., et al. (2021). Structural variants selected during yak domestication inferred from long-read whole-genome sequencing. *Mol. Biol. Evol.* 38 (9), 3676–3680. doi:10.1093/molbev/msab134
- Zhang, T., Wang, T., Niu, Q., Zheng, X., Li, H., Gao, X., et al. (2022). Comparative transcriptomic analysis reveals region-specific expression patterns in different beef cuts. *BMC Genomics* 23 (1), 387. doi:10.1186/s12864-022-08527-3
- Zhang, X., Wang, K., Wang, L., Yang, Y., Ni, Z., Xie, X., et al. (2016). Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics* 17, 379. doi:10.1186/s12864-016-2702-6



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Liu Yang,
Agricultural Research Service (USDA),
United States
Siyuan Zhan,
Sichuan Agricultural University, China

*CORRESPONDENCE

Jianmin Wang
✉ wangjm@sdaa.edu.cn

RECEIVED 21 March 2024

ACCEPTED 30 May 2024

PUBLISHED 12 June 2024

CITATION

Li Q, Chao T, Wang Y, Xuan R, Guo Y, He P,
Zhang L and Wang J (2024) Transcriptome
analysis revealed the characteristics and
functions of long non-coding RNAs in the
hypothalamus during sexual maturation in
goats.

Front. Vet. Sci. 11:1404681.

doi: 10.3389/fvets.2024.1404681

COPYRIGHT

© 2024 Li, Chao, Wang, Xuan, Guo, He,
Zhang 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.

Transcriptome analysis revealed the characteristics and functions of long non-coding RNAs in the hypothalamus during sexual maturation in goats

Qing Li^{1,2}, Tianle Chao^{1,2}, Yanyan Wang^{1,2}, Rong Xuan^{1,2},
Yanfei Guo^{1,2}, Peipei He^{1,2}, Lu Zhang^{1,2} and Jianmin Wang^{1,2*}

¹Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, College of Animal Science and Veterinary Medicine, Shandong Agricultural University, Tai'an, China,

²Key Laboratory of Efficient Utilization of Non-Grain Feed Resources (Co-Construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Shandong Agricultural University, Tai'an, China

The hypothalamus is an essential neuroendocrine area in animals that regulates sexual development. Long non-coding RNAs (lncRNAs) are hypothesized to regulate physiological processes related to animal reproduction. However, the regulatory mechanism by which lncRNAs participate in sexual maturity in goats is poorly known, particularly from birth to sexual maturation. In this study, RNAseq analysis was conducted on the hypothalamus of four developmental stages (1day (D1, $n = 5$), 2 months (M2, $n = 5$), 4months (M4, $n = 5$), and 6 months (M6, $n = 5$)) of Jining grey goats. The results showed that a total of 237 differentially expressed lncRNAs (DELs) were identified in the hypothalamus. Among these, 221 DELs exhibited cis-regulatory effects on 693 target genes, while 24 DELs demonstrated trans-regulatory effects on 63 target genes. The target genes of these DELs are mainly involved in biological processes related to energy metabolism, signal transduction and hormone secretion, such as sphingolipid signaling pathway, adipocytokine signaling pathway, neurotrophic signaling pathway, glutamatergic synapse, P53 signaling pathway and GnRH signaling pathway. In addition, XR_001918477.1, TCONS_00077463, XR_001918760.1, and TCONS_00029048 and their potential target genes may play a crucial role in the process of goat sexual maturation. This study advances our understanding of lncRNA in hypothalamic tissue during sexual maturation in goats and will give a theoretical foundation for improving goat reproductive features.

KEYWORDS

goat, long non-coding RNA, hypothalamus, transcriptome, sexual maturation

1 Introduction

The sexual maturity of animals strongly influences their reproductive capacity. Animals go through puberty after birth to reach sexual maturity, a process involving complex physiological changes (1). Early-maturing ruminants experience a younger age of first birth, leading to enhanced reproductive longevity and fecundity (2). The hypothalamus, the gonadal axis' most upstream tissue and organ, secretes GnRH to stimulate the synthesis of pituitary gonadotropins and gonadal steroid hormones, which are vital for an animal's sexual

development (3–5). The hypothalamus receives signals from the periphery and others that act directly or indirectly on GnRH and its associated reproductive neurons, which in turn affect the synthesis, secretion, and morphology of GnRH, ultimately leading to the occurrence of puberty and sexual maturation (6–8).

Long non-coding RNAs (lncRNAs) are a prevalent class of non-coding RNAs found in mammals, typically exceeding 200 nt in length and devoid of protein-coding capabilities (9, 10). It has been demonstrated to play a significant role in diverse biological processes through transcriptional, post-transcriptional, or epigenetic regulation (11). Currently, lncRNAs have been widely reported to be involved in embryonic development (12), muscle development (13), metabolism (14), and reproductive regulation (15). Recent research has demonstrated that lncRNA Meg3 can regulate the expression of GnRH and Kiss-1 in hypothalamic cells, and knockdown of lncRNA Meg3 can delay puberty in female rats (16). Additionally, the lncRNA MSTRG.33887.2 has the potential to influence goat reproduction by regulating target genes involved in hypothalamic folate metabolism and energy metabolism homeostasis (17). Mouse hypothalamic lncRNA AK044061 plays a crucial role in energy balance by mediating NF- κ B. Neurons with high expression of lncRNAs AK044061 in ARC cells lead to energy imbalance and obesity in mice (18).

However, there is a scarcity of research investigating the dynamic expression patterns of lncRNAs throughout the sexual maturation process in female goats, specifically from birth to the completion of sexual maturation. The Jining grey goat, a well-known high-breeding goat breed in China, exhibits non-seasonal estrus, strong fecundity, and precocious puberty. Sexual maturity in these goats is reached at 3 to 4 months of age, with puberty commencing as early as 2 months of age (19). This trait makes it an ideal animal model for investigating goat fecundity. Hence, it is highly significant to investigate the regulatory mechanism of hypothalamic lncRNAs in the sexual maturation process of female goats.

In this study, the lncRNA of hypothalamic tissue of 1-day-old, 2-month-old, 4-month-old, and 6-month-old (D1, M2, M4, and M6; $n = 5$) female Jining grey goats were sequenced. This study aimed to uncover the expression profile characteristics of lncRNAs during goat sexual maturation, identify lncRNAs associated with hypothalamic development and sexual maturation, and elucidate their molecular regulatory mechanisms. Our research will provide a theoretical basis for the genetic improvement of goat reproductive traits.

2 Materials and methods

2.1 Animals and sample collection

The experimental goats were all from the Jining Grey Goats Breeding Farm (Jiaxiang, Shandong, China). Under the same feeding management conditions, 20 healthy and disease-free female Jining grey goats were selected. The selected goats were divided into four groups according to age: 1 day old (D1, $n = 5$; body weight (BW): 2.08 ± 0.11 kg), 2 months old (M2, $n = 5$; BW: 4.42 ± 0.24 kg), 4 months old (M4, $n = 5$; BW: 7.62 ± 0.50 kg), and 6 months of age (M6, $n = 5$; BW: 8.82 ± 0.53 kg). The body condition of Jining grey goats was similar in each group. The experimental goats were slaughtered on the same day, and after the electric shock, the hypothalamic tissue was quickly slaughtered and collected, and stored at -80°C .

2.2 RNA extraction and library construction

Total RNA was extracted from 20 hypothalamic tissues using TRIzol® reagent (Thermo Fisher Scientific, Waltham, MA, United States). Screening of qualified RNA samples for RNA strand-specific library construction. The rRNA was removed from total RNA samples using the Ribo-Zero rRNA Removal Kit (Illumina, Inc., San Diego, United States), and then a sequencing library was generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB E7420) for Illumina to generate sequencing libraries. The enriched RNA was fragmented using a fragmentation buffer to yield small fragments. Then, the fragmented RNA served as a template for reverse transcription with the addition of 6bp random primers (random hexamers) to synthesize the first cDNA strand. This was followed by the addition of buffer, dNTPs (with dTTP replaced by dUTP), DNA polymerase I, and RNase H to synthesize the second cDNA strand. The synthesized double-stranded cDNA was purified and enriched via PCR. The PCR product was then purified to obtain the final strand-specific library. After reverse transcription and PCR amplification, 150bp paired-end reads were sequenced using the Illumina Novaseq6000 platform (Illumina, Inc., San Diego, United States).

2.3 Reads mapping and transcriptome assembly

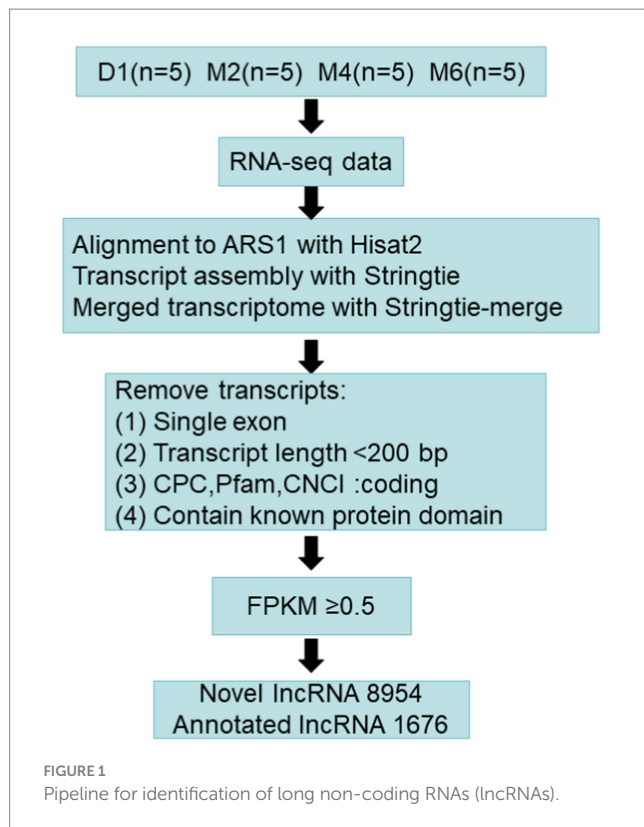
To obtain high-quality sequencing data, we utilized Fastp (v0.23.1) to eliminate sequences containing poly-N, low-quality reads, and adapters from the obtained sequencing data. The high-quality reads obtained are used for downstream data analysis. We generated an index of the reference genome by employing HISAT2 (v2.0.5.) Subsequently, we aligned the clean reads with the goat reference genome (GCF_001704415.2_ARSL1.2) using HISAT2 (20). Transcript assembly is performed using Stringtie (v1.3.3b), and gene expression levels are calculated (21). Gene expression levels were normalized using fragments per kilobase of exon model per million mapped reads (FPKM).

2.4 lncRNA identification

The novel lncRNAs were identified in hypothalamic tissue following the steps shown in Figure 1: (1) removing transcripts with an exon number of 1, (2) removing transcripts less than 200 nt in length, and (3) screening out transcripts that overlapped the exon region annotated in the database by gffcompare software (v0.10.6) (22); (4) CPC2 (v3.2.0) (23), Pfam (v1.6) (24), and CNCI (v2.0) (25) were used to predict the encoding potential of lncRNAs, and the transcripts that were predicted in the three software without coding potential were intersected, and (5) the low-expression lncRNAs (FPKM < 0.5) were filtered out.

2.5 Differential expression analysis of lncRNAs

The DESeq2 (v1.20.0) package was utilized to examine the differential expression of lncRNAs (DElNs) across various



developmental stages. The readcounts from the sequencing data were used as the input matrix, and the p -value was adjusted utilizing the Benjamini & Hochberg method. The DELs were screened according to the threshold $|\log_2(\text{Fold change})| \geq 1$ and False Discovery Rate (FDR) < 0.05 . Cluster analysis of FPKM values of lncRNAs was performed using the ggplot2 package (v3.4.4). Data for lncRNAs were normalized $[\log_2(X + 1)]$ and then standardization (z -score).

2.6 Prediction and functional analysis of potential target genes of lncRNA

lncRNAs can regulate the expression of potential target genes through cis- or trans-regulatory methods. Based on the location information of lncRNAs, mRNAs within 100 kb upstream and downstream of lncRNAs are defined as cis-target genes of lncRNAs (26). There will be a significant correlation with lncRNA expression ($|R| > 0.95$ and $p < 0.05$) is defined as a potential trans-target gene for lncRNA.

Subsequently, we used clusterProfiler software (v3.8.1) to perform Gene Ontology (GO) functional analysis of these differentially lncRNAs predicted target genes (27). The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis is performed by KOBAS (<http://bioinfo.org/kobas>; Accessed: 12.19, 2023) (28). $p < 0.05$ was considered to be significantly enriched.

2.7 Quantitative real-time PCR

Six lncRNAs were randomly selected and the accuracy of the lncRNA sequencing results was verified by qRT-PCR. First, we used

PrimeScript™II First strand cDNA synthesis kit (Takara Bio Inc., Dalian) to reverse transcribe total RNA from goat hypothalamus tissue into cDNA. Then qRT-PCR was performed in a Roche LightCycler 96 using the SYBR PrimeScript™ RT-PCR Kit (Takara Bio Inc., Dalian). GAPDH was used as an internal reference gene to correct gene expression levels and normalize the data. The primers designed using Primer 5.0 (Supplementary Table S1). The relative expression levels of lncRNAs were calculated by $2^{-\Delta\Delta CT}$ (29). One-way ANOVA was performed with SPSS 17.0, and the results were expressed as mean $\pm$ standard error. Three repetitions are performed for each set. $p < 0.05$ was considered statistically significant.

3 Result

3.1 Overview of RNA sequencing data

RNA was isolated from the hypothalamic tissues of female Jining grey goats at four developmental stages (D1, M2, M4, and M6), and 20 lncRNA libraries were constructed. Sequencing of the libraries was conducted on the Novaseq 6000 platform, resulting in a total of 1,803,310,692 raw reads. Following quality control procedures, we obtained 1,764,056,912 clean reads (Supplementary Table S2). The alignment of these clean reads to the goat reference genome was performed using HiSAT2, achieving alignment rates ranging from 86.05 to 95.99%, with unique mapping reads alignment rates between 75.1 and 91.98% (Supplementary Table S2). Subsequent analysis only considered the uniquely mapped reads.

3.2 Identification and characterization of lncRNAs

A total of 10,630 lncRNAs were identified according to the steps shown in Figure 1, of which 1,676 were annotated and 8,954 lncRNAs were newly identified. Cluster analysis showed that most of the lncRNAs were expressed at low levels at the D1 stage (Supplementary Figure S1). Further analysis of the identified lncRNA signatures showed that about 62% of the lncRNAs had 2 exons, and a few lncRNAs (3%) had more than 6 exons (Figure 2A). In addition, the length distribution of the identified lncRNAs ranged from 122 to 73,429 bp. More than 50% of the lncRNAs were less than 1,000 bp in length, about 83% were in the 0–3 kb range, and a few (17%) were greater than 3 kb (Figure 2B). About 35.2% of the lncRNAs are located in the intergenic region. Only 12.4% of the lncRNAs are from the antisense region, and about 32.9% of the lncRNAs are from the intron region (Figure 2C).

3.3 Differential expression analysis of lncRNAs

According to the screening criteria of $|\log_2(\text{Fold change})| \geq 1$ and $p_{\text{adj}} < 0.05$, the lncRNAs of four different developmental stages (D1, M2, M4 and M6) in hypothalamic tissue were compared and analysed. This analysis resulted in the identification of 237 differentially expressed lncRNAs (Figure 3A).

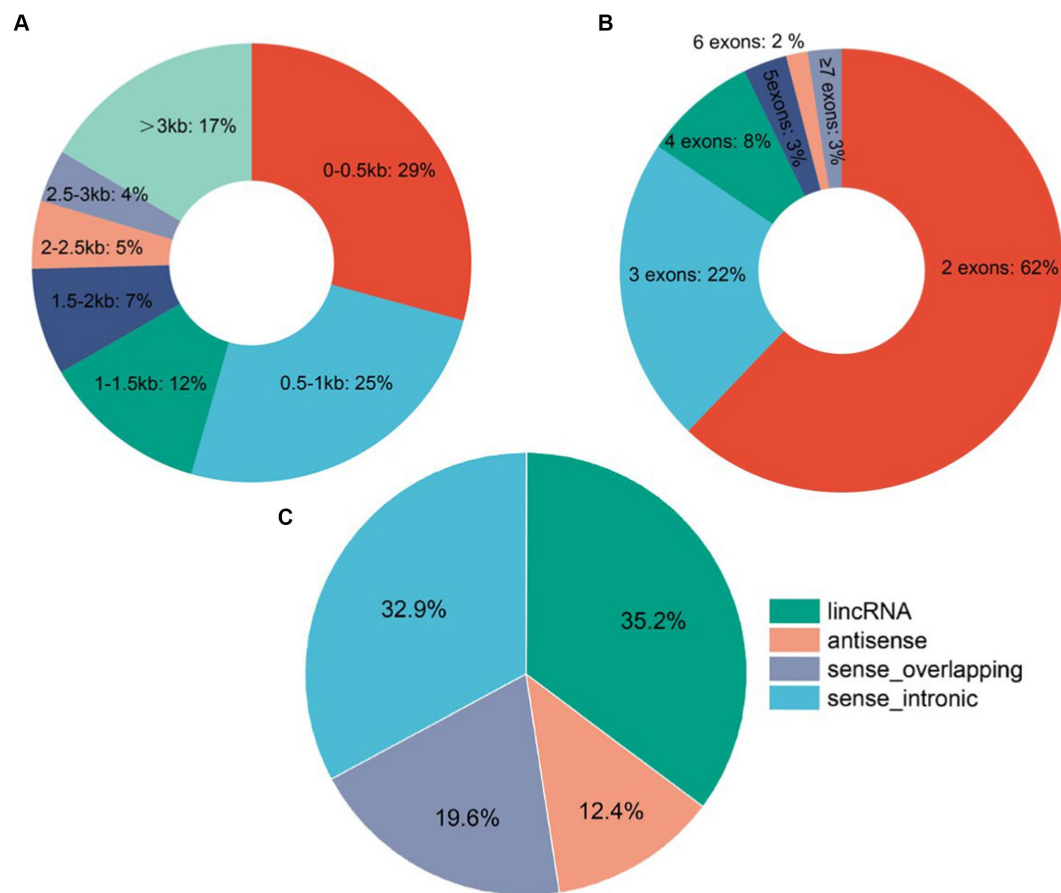


FIGURE 2 Identification and characterization of goat hypothalamic lncRNAs. **(A)** Pie plot of lncRNA exon number distribution. **(B)** Pie plot of lncRNA length distribution. **(C)** Pie chart of lncRNA classification.

A total of 135 DELs (31 up-regulated and 104 down-regulated) were identified in M2 vs. D1, 68 DELs (38 up-regulated and 30 down-regulated) were identified in M4 vs. M2, and 16 DELs (11 up-regulated and 5 down-regulated) were identified in M6 vs. M4. A total of 53 DELs (16 up-regulated and 37 down-regulated) were identified in M4 vs. D1, 33 DELs (15 up-regulated and 18 down-regulated) were identified in M6 vs. D1, and 69 DELs (48 up-regulated and 21 down-regulated) were identified in M6 vs. M2. Interestingly, Figure 3B shows different expression patterns for these DELs, and Figure 3C shows that the M2 vs. D1 group has the most unique DELs, with 72. The M6 vs. M4 group had the lowest number of DELs with 3. Remarkably, 9 DELs were the same in the three comparison groups of M2 vs. D1, M4 vs. D1, and M6 vs. D1. This suggests that TCONS_00076225, TCONS_00148767, TCONS_00191611, TCONS_00102342, TCONS_00192735, TCONS_00032355, TCONS_00032357, TCONS_00053700, and TCONS_00070238 may be significant in the postnatal sexual development of goats.

3.4 lncRNA target gene prediction and functional analysis

lncRNAs have been implicated in influencing gene expression through cis-or trans-interactions. To investigate the role of lncRNAs

in the goat hypothalamus, we predicted the potential regulatory roles of the identified DELs on both cis and trans target genes. Specifically, based on a distance threshold of 100 kb between lncRNAs and their target genes, 221 DELs were predicted to regulate 693 target genes in a cis manner (Supplementary Table S3).

The Gene Ontology (GO) analysis results showed that the target genes were significantly enriched in 50 categories (Supplementary Table S4). These cis-target genes are involved in many biological processes, such as regulation of catalytic activity, regulation of molecular function, regulation of hydrolase activity, and dephosphorylation (Figure 4A). In addition, KEGG analysis showed that these target genes were significantly enriched in sphingolipid signaling pathway, protein processing in endoplasmic reticulum, progesterone-mediated oocyte maturation, adipocytokine signaling pathway, neurotrophin signaling pathway, oocyte meiosis, sphingolipid metabolism, glutamatergic synapse, cGMP-PKG signaling pathway, inositol phosphate metabolism, phospholipase D signaling pathway and other pathways ($p < 0.05$) (Figure 4B and Supplementary Table S5).

According to $|R| > 0.95$ and $p < 0.05$, 24 lncRNAs were found, which had 63 trans target genes (Supplementary Table S6 and Supplementary Figure S2). The number of trans-target genes identified in lncRNAs was significantly lower than that in cis, suggesting that lncRNAs may function mainly by cis-regulating

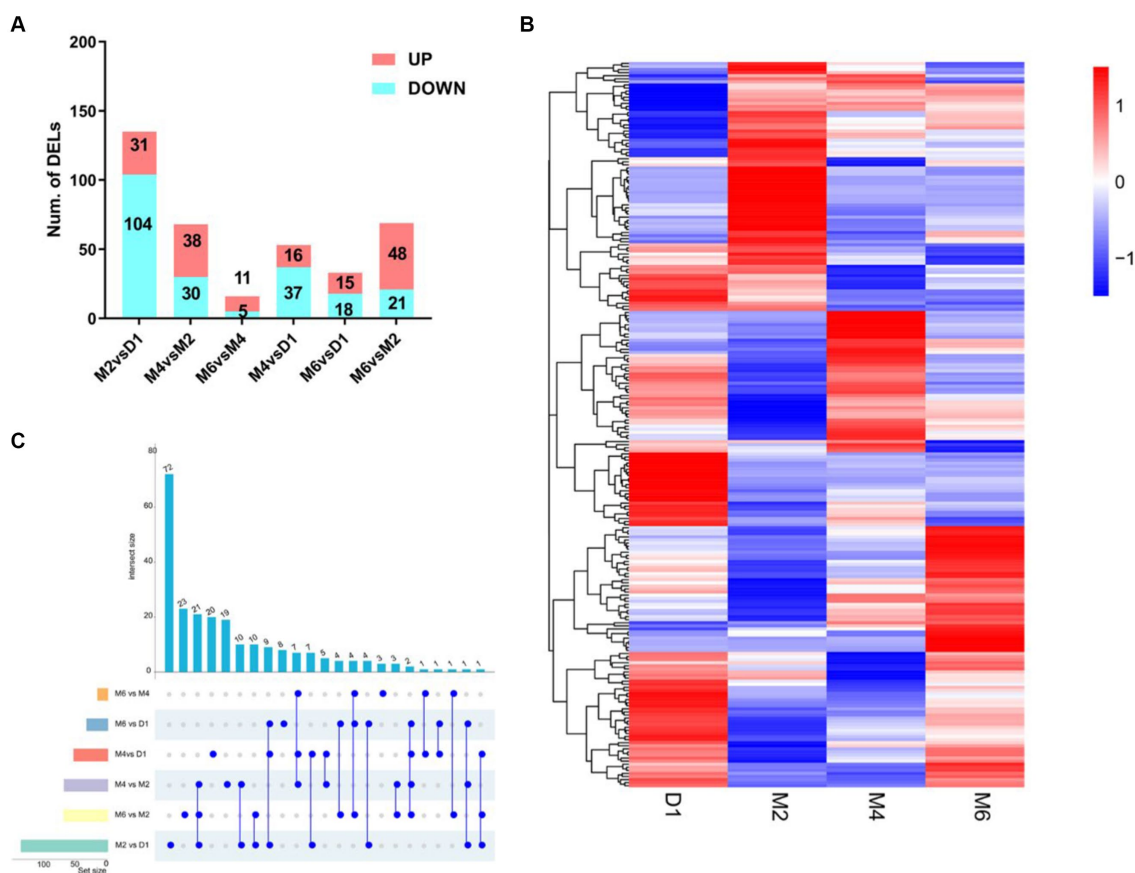


FIGURE 3

Differentially expressed lncRNA (DELs) characteristics in different comparison groups. **(A)** Cluster analysis of differential lncRNAs. Each column represents a grouping and each row represents a differential lncRNA. **(B)** Histogram of DELs. **(C)** The upset plot shows the distribution of DELs in the different comparison groups.

gene expression. The GO analysis results showed that we found that these cis-target genes were involved in many biological processes, such as signal receptor binding, hormone activity, NADH dehydrogenase activity, transferase activity, transfer aminoacyl, etc. (Figure 4C and Supplementary Table S7). In addition, KEGG analysis showed that these target genes were significantly enriched in neuroactive ligand receptors, multiple apoptosis, adipocytokine signaling pathway, GnRH signaling pathway, JAK-STAT signaling pathway, and p53 signaling pathway ($p < 0.05$) (Figure 4D and Supplementary Table S8).

The results suggest that these lncRNAs may be involved in hormone secretion, signal transduction processes, and thus regulation of sexual maturation in goats by modulating cis and trans target genes. Interestingly, LOC108634846, LOC108635405, LOC108637396, AGRP, and PIK3C2G were predicted in both cis and trans target genes. Among these, PIK3C2G is the trans-target gene and the cis-target gene of TCONS_00038560.

3.5 Interaction analysis

To better understand the role of goat hypothalamic DELs in the process of sexual maturation, we selected the target gene of DELs involved in reproduction. Construction of lncRNA-mRNA regulatory

networks for DELs and their cis- and trans-target genes, respectively (Figure 5). A total of 31 lncRNAs regulated 34 target genes.

3.6 Verify lncRNA expression using qRT-PCR

To verify the accuracy of the sequencing results, we randomly selected 6 lncRNAs (XR_001297374.2, XR_001917241.1, TCONS_00176496, XR_001919854.1, TCONS_00032357, and TCONS_00074891) for qRT-PCR detection. The results showed that the expression patterns of these lncRNAs and those found in the transcriptome data were consistent with the sequencing results (Figure 6). This further illustrates the high reliability and accuracy of RNA-seq.

4 Discussion

The hypothalamus is an important neuroendocrine center in mammals and plays an important role in the sexual development of animals. lncRNAs are non-coding RNAs that are more than 200 nt in length and do not have protein-coding functions. Many studies have shown that lncRNAs play an important role in reproductive regulation

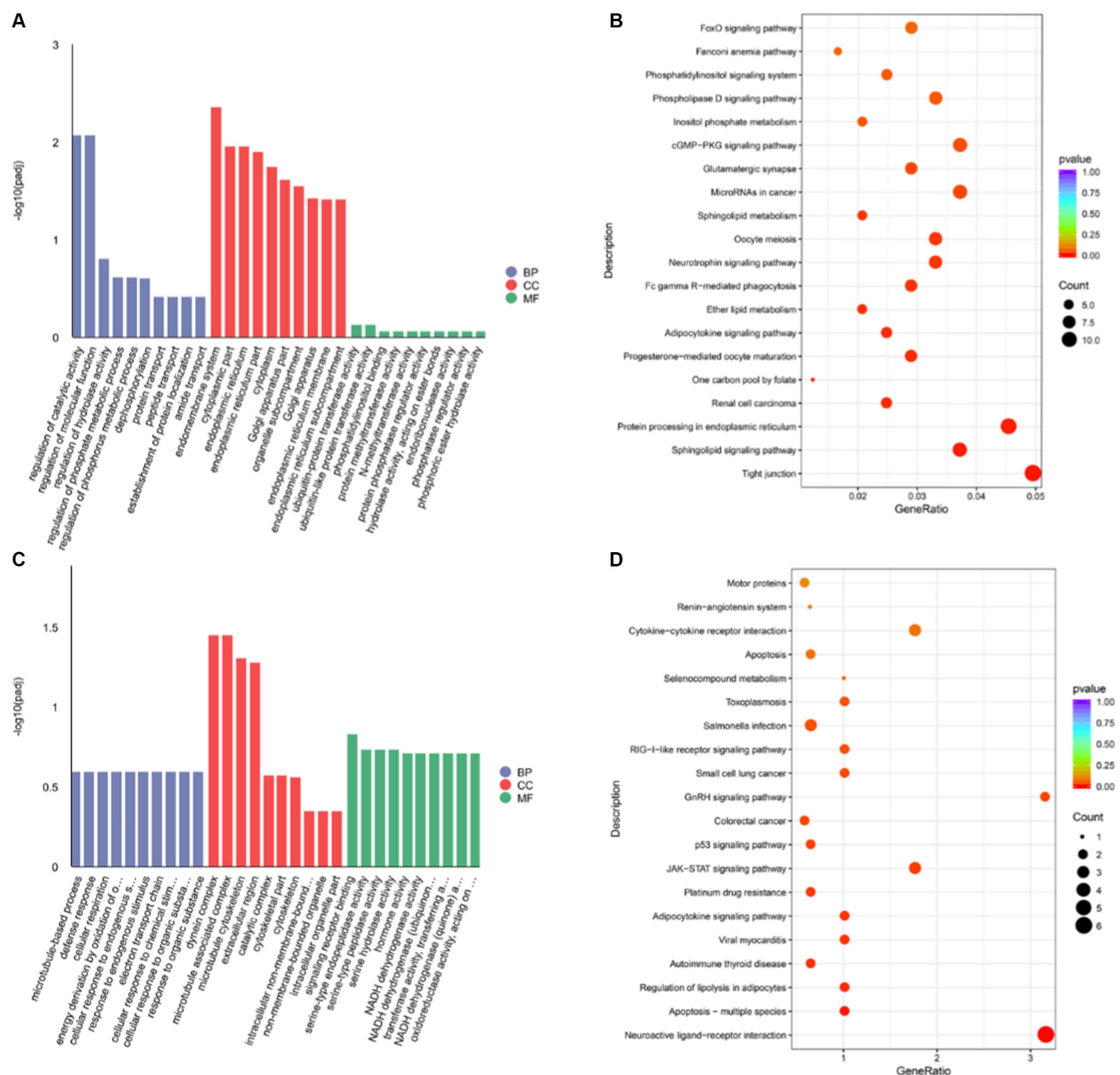


FIGURE 4

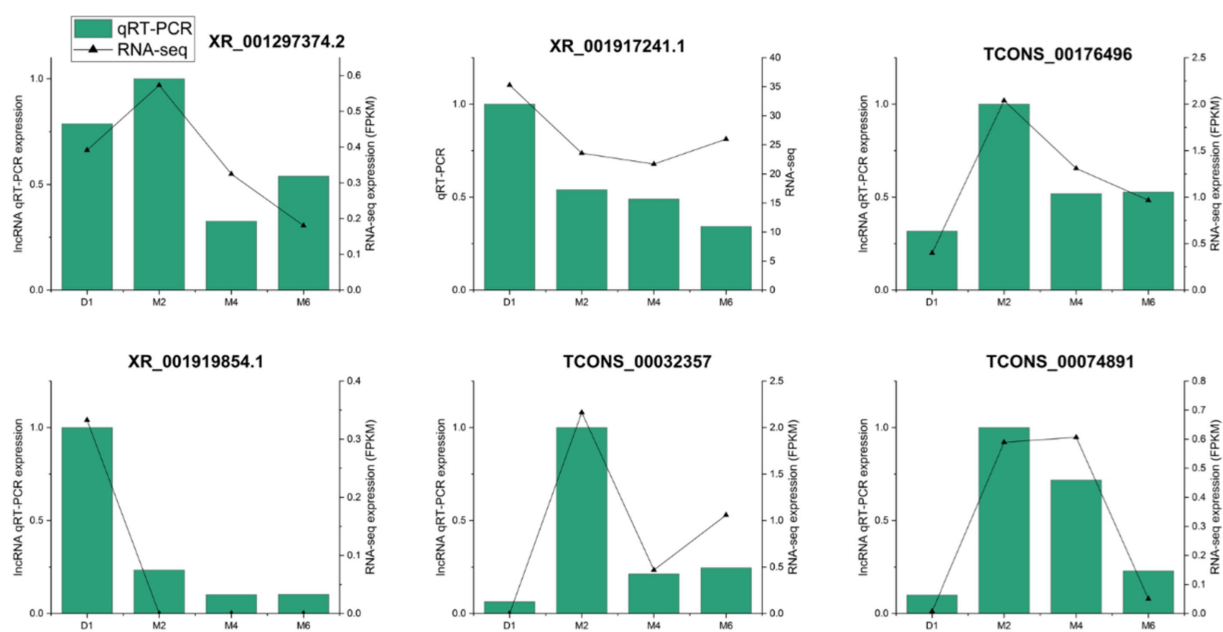
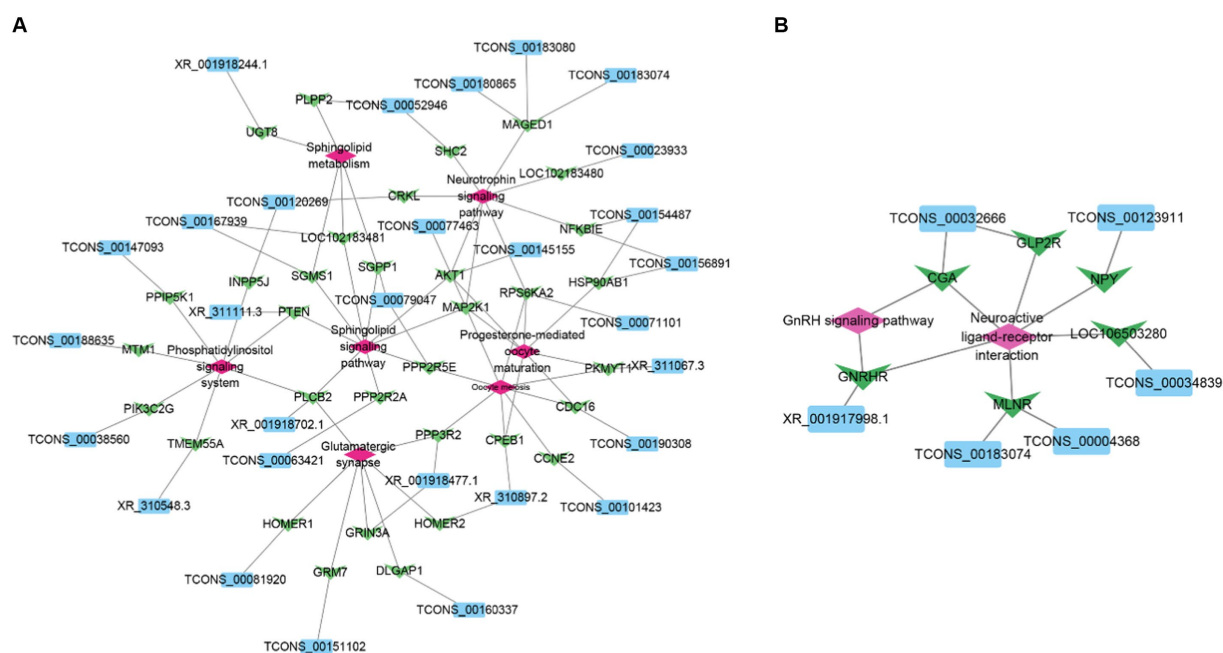
The functional enrichment analysis of DELs target genes (A) DELs cis-target gene GO analysis. (B) DELs cis-target gene KEGG analysis. (C) DELs trans-target gene GO analysis. (D) DELs trans-target gene KEGG analysis.

through transcriptional regulation or epigenetic modification (30, 31). Investigating the role of hypothalamic lncRNAs during sexual maturation in female goats is critical to understanding reproductive mechanisms in this species.

This study conducted lncRNA sequencing on hypothalamic tissue samples obtained from four postnatal developmental stages (D1, M2, M4, and M6) of female jining grey goats. A total of 10,254 lncRNAs were identified according to rigorous screening criteria, including 1,676 known lncRNAs and 8,578 novel identified lncRNAs. Characterization of lncRNAs revealed that the majority of them consisted of 2–3 exons (84%), and approximately 54% were under 1 kb in length, consistent with previously reported profiles of lncRNAs in goats (32). A total of 236 DELs were identified by differential analysis of lncRNAs. The number of DELs is highest between M2 and D1, totaling 135, while the DELs are least frequent between M6 and M4, with only 16 instances. These results

indicated that the number of DELs decreased with the gradual maturity of goats, and the development of jining grey goats from birth to 2 months of age was an important developmental stage.

Recent studies have shown that lncRNAs are involved in the regulation of gene expression through both cis- and trans-regulatory mechanisms and play an important role in a wide range of biological processes (33, 34). The results showed that 220 DELs regulated 693 target genes through cis. These target genes are involved in the sphingomyelin signaling pathway, progesterone-mediated oocyte maturation, neurotrophin signaling pathway, oocyte meiosis, sphingomyelin metabolism, glutamatergic synapses, cGMP-PKG signaling pathway, phosphoinositide metabolism, and phospholipase D signaling pathway. These pathways are involved in hypothalamic neuronal development, energy metabolism, and reproductive processes (35–38), suggesting that lncRNAs may play an important role in the



superfamily, and the expression level of GRIN3A is significantly increased before estrus in mice, which may be related to enhanced glutamate receptor signaling in preovulatory GnRH neurons (43). In mammals, LEP acts on neural circuits in the hypothalamus to regulate feeding and energy metabolism (44), and in addition, leptin can be involved in the regulation of puberty through mTOR (9).

The trans target genes of lncRNAs are predicted by calculating the correlation of mRNA to lncRNAs. This approach enables lncRNAs to regulate mRNA away from their transcription sites (9). In this study, a total of 23 lncRNAs were negatively correlated with 63 protein-coding genes. These target genes are mainly enriched in neuroactive ligand and receptor interactions, JAK-STAT signaling pathway, p53 signaling pathway, and GnRH signaling pathway. Based on this, we speculate that these lncRNAs may be involved in various biological processes, such as hormone secretion and signal transduction, during goat sexual maturation by regulating the expression of these target genes (45–47). Furthermore, we constructed a regulatory network related to reproduction, including 6 lncRNAs and 6 target genes involved in the interaction between neuroactive ligands and receptors and the GnRH signaling pathway. Among them, TCONS_00123911 acts on neuropeptide Y (NPY) through trans regulation. Studies have shown that NPY plays an important role in energy homeostasis and reproductive hormone secretion (48). In addition, NPY has been shown to play an important role in sexual maturation by regulating GnRH secretion patterns and luteinizing hormone secretion (49, 50). In addition, we have identified XR_001917998.1 that can trans-modulate GnRHR. GnRHR plays an important role in the regulation of mammalian reproduction (51). Studies have shown that after hypothalamic GnRH stimulation, GnRHR regulates the activity of the HPG axis through signal transduction, thereby participating in the synthesis and release of LH and FSH, and regulating gonadal function (52, 53).

The identified DELs potentially play a role in the sexual maturation process of goats by modulating target genes through cis-trans regulation. However, experimental validation is required to confirm the functions of lncRNAs. Subsequent research will focus on unraveling the molecular mechanisms through which lncRNAs regulate sexual maturation in goats at both the molecular and cellular levels.

5 Conclusion

In summary, we described the characteristics of the expression profile of lncRNA in the hypothalamus at four developmental stages in goats and analyzed the regulatory mechanism of lncRNAs in the process of sexual maturation in goats. In this study, a total of 237 DELs were identified and their cis-trans target genes were predicted, and functional analysis showed that the cis-trans target genes of these DELs were mainly involved in sphingomyelin signaling pathway, glutamatergic synapse, neuroactive ligand and receptor interaction, p53 signaling pathway, GnRH signaling pathway, hypothalamic development and hormone secretion. This work enriches the goat lncRNA database, lays a theoretical foundation for elucidating the molecular mechanism of goat sexual maturation in the future, and will provide a theoretical basis for the improvement of goat genetic traits.

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 at: <https://www.ncbi.nlm.nih.gov/GSE244004>.

Ethics statement

The animal study was approved by the Animal Care and Use Committee of Shandong Agricultural University ethics committee (SDAUA-2023-157). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

QL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. TC: Funding acquisition, Validation, Visualization, Writing – review & editing. YW: Software, Writing – review & editing. RX: Validation, Visualization, Writing – review & editing. YG: Visualization, Writing – review & editing. PH: Software, Writing – review & editing. LZ: Visualization, Writing – review & editing. JW: Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This project was financed by the Project of Improved Agricultural Varieties in Shandong Province (2021LZGC010 and 2019LZGC012), the National Key R & D Plan (Integration and Demonstration of Key Technologies for Large-Scale, Efficient and Safe Breeding of Goats) (2018YFD0501906), Natural Science Foundation Project of Shandong Province (ZR2023QC044), Shandong Provincial Sheep and Goat Industry Technology System (SDAIT-10-01), Shandong “Double Tops” Program (SYL2017YSTD12).

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.2024.1404681/full#supplementary-material>

References

- DeLamater J, Friedrich WN. Human sexual development. *J Sex Res.* (2002) 39:10–4. doi: 10.1080/00224490209552113
- Valasi I, Chadio S, Fthenakis GC, Amiridis GS. Management of pre-pubertal small ruminants: physiological basis and clinical approach. *Anim Reprod Sci.* (2012) 130:126–34. doi: 10.1016/j.anireprosci.2012.01.005
- Brown GR, Spencer KA. Steroid hormones, stress and the adolescent brain: a comparative perspective. *Neuroscience.* (2013) 249:115–28. doi: 10.1016/j.neuroscience.2012.12.016
- Herbison AE. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. *Nat Rev Endocrinol.* (2016) 12:452–66. doi: 10.1038/nrendo.2016.70
- Tenuta M, Carlomagno F, Cangiano B, Kanakis G, Pozza C, Sbardella E, et al. Somatotrophic-testicular axis: a crosstalk between GH/IGF-I and gonadal hormones during development, transition, and adult age. *Andrology.* (2021) 9:168–84. doi: 10.1111/andr.12918
- Dees WL, Hiney JK, Srivastava VK. IGF-1 influences gonadotropin-releasing hormone regulation of puberty. *Neuroendocrinology.* (2021) 111:1151–63. doi: 10.1159/000514217
- Pellegrino G, Martin M, Allet C, Lhomme T, Geller S, Franssen D, et al. GnRH neurons recruit astrocytes in infancy to facilitate network integration and sexual maturation. *Nat Neurosci.* (2021) 24:1660–72. doi: 10.1038/s41593-021-00960-z
- Spaziani M, Tarantino C, Tahani N, Gianfrilli D, Sbardella E, Lenzi A, et al. Hypothalamo-pituitary axis and puberty. *Mol Cell Endocrinol.* (2021) 520:111094. doi: 10.1016/j.mce.2020.111094
- Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* (2018) 172:393–407. doi: 10.1016/j.cell.2018.01.011
- Bridges MC, Daulagala AC, Kourtidis A. LNCcation: lnc RNA localization and function. *J Cell Biol.* (2021) 220:e202009045. doi: 10.1083/jcb.202009045
- Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol.* (2013) 10:925–33. doi: 10.4161/rna.24604
- Zhang Y, Duan E. lncRNAs and paraspeckles predict cell fate in early mouse embryo. *Biol Reprod.* (2019) 100:1129–31. doi: 10.1093/biolre/iox021
- Maguire EM, Xiao Q. Noncoding RNAs in vascular smooth muscle cell function and neointimal hyperplasia. *FEBS J.* (2020) 287:5260–83. doi: 10.1111/febs.15357
- Shen X, Zhang Y, Ji X, Li B, Wang Y, Huang Y, et al. Long noncoding RNA lncRHL regulates hepatic VLDL secretion by modulating hnRNPU/BMAL1/MTTP axis. *Diabetes.* (2022) 71:1915–28. doi: 10.2337/db21-1145
- Liu KS, Li TP, Ton H, Mao XD, Chen YJ. Advances of long noncoding RNAs-mediated regulation in reproduction. *Chin Med J.* (2018) 131:226–34. doi: 10.4103/0366-6999.222337
- Li H, Ren C, Lu J, Xu S, Gong X, Zhang W, et al. Knockdown of lncRNA Meg3 delays the onset of puberty in female rats. *Theriogenology.* (2023) 207:72–81. doi: 10.1016/j.theriogenology.2023.05.026
- Dong S, Hou B, Yang C, Li Y, Sun B, Guo Y, et al. Comparative hypothalamic transcriptome analysis reveals crucial mRNAs, lncRNAs, and circRNAs affecting litter size in goats. *Genes.* (2023) 14:444. doi: 10.3390/genes14020444
- Li J, Long J, Zhang Q, Shen H, Guo AY, Ma Z, et al. Hypothalamic long noncoding RNA AK044061 is involved in the development of dietary obesity in mice. *Int J Obes.* (2021) 45:2638–47. doi: 10.1038/s41366-021-00945-9
- Shi Y, Wang S, Bai S, Huang L, Hou Y. Postnatal ovarian development and its relationship with steroid hormone receptors in Jining Grey goats. *Anim Reprod Sci.* (2015) 154:39–47. doi: 10.1016/j.anireprosci.2015.01.001
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* (2019) 37:907–15. doi: 10.1038/s41587-019-0201-4
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290–5. doi: 10.1038/nbt.3122
- Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Res.* (2020) 9:304. doi: 10.12688/f1000research.23297.1
- Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* (2017) 45:W12–w16. doi: 10.1093/nar/gkx428
- Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* (2014) 42:D222–30. doi: 10.1093/nar/gkt1223
- 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
- Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. lncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* (2019) 47:D1034–7. doi: 10.1093/nar/gky905
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* (2011) 39:W316–22. doi: 10.1093/nar/gkr483
- Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^{-ΔΔC_T} method for quantitative real-time polymerase chain reaction data analysis. *Biostat Bioinforma Biomath.* (2013) 3:71–85.
- Taylor DH, Chu ET, Spektor R, Soloway PD. Long non-coding RNA regulation of reproduction and development. *Mol Reprod Dev.* (2015) 82:932–56. doi: 10.1002/mrd.22581
- Golicz AA, Bhalla PL, Singh MB. lncRNAs in plant and animal sexual reproduction. *Trends Plant Sci.* (2018) 23:195–205. doi: 10.1016/j.tplants.2017.12.009
- Varela-Martínez E, Luigi-Sierra MG, Guan D, López-Béjar M, Casas E, Olvera-Maneu S, et al. The landscape of long non-coding RNA expression in the goat brain. *J Dairy Sci.* (2024) 107:4075–91. doi: 10.3168/jds.2023-23966
- Roberts TC, Morris KV, Weinberg MS. Perspectives on the mechanism of transcriptional regulation by long non-coding RNAs. *Epigenetics.* (2014) 9:13–20. doi: 10.4161/epi.26700
- Amaral PP, Leonardi T, Han N, Viré E, Gascoigne DK, Arias-Carrasco R, et al. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* (2018) 19:32. doi: 10.1186/s13059-018-1405-5
- Stoffel W, Jenke B, Blöck B, Zumbansen M, Koebke J. Neutral sphingomyelinase 2 (smpd3) in the control of postnatal growth and development. *Proc Natl Acad Sci USA.* (2005) 102:4554–9. doi: 10.1073/pnas.0406380102
- Vozella V, Realini N, Misto A, Piomelli D. Feeding stimulates sphingosine-1-phosphate mobilization in mouse hypothalamus. *Int J Mol Sci.* (2019) 20:4008. doi: 10.3390/ijms20164008
- Heras V, Castellano JM, Fernandois D, Velasco I, Rodríguez-Vazquez E, Roa J, et al. Central ceramide signaling mediates obesity-induced precocious puberty. *Cell Metab.* (2020) 32:951–966.e958. doi: 10.1016/j.cmet.2020.10.001
- Gueorguiev M, Góth ML, Korbonits M. Leptin and puberty: a review. *Pituitary.* (2001) 4:79–86. doi: 10.1023/a:1012943029127
- Burger LL, Haisenleder DJ, Aylor KW, Marshall JC. Regulation of Lhb and Egr 1 gene expression by GnRH pulses in rat pituitaries is both c-Jun N-terminal kinase (JNK)- and extracellular signal-regulated kinase (ERK)-dependent. *Biol Reprod.* (2009) 81:1206–15. doi: 10.1095/biolreprod.109.079426
- Batchuluun K, Azuma M, Yashiro T, Kikuchi M. Notch signaling-mediated cell-to-cell interaction is dependent on E-cadherin adhesion in adult rat anterior pituitary. *Cell Tissue Res.* (2017) 368:125–33. doi: 10.1007/s00441-016-2540-5
- Bossi S, Pizzamiglio L, Paoletti P. Excitatory GluN1/GluN3A glycine receptors (eGlyRs) in brain signaling. *Trends Neurosci.* (2023) 46:667–81. doi: 10.1016/j.tins.2023.05.002
- Vastagh C, Rodolose A, Solymosi N, Liposits Z. Altered expression of genes encoding neurotransmitter receptors in GnRH neurons of proestrous mice. *Front Cell Neurosci.* (2016) 10:230. doi: 10.3389/fncel.2016.00230
- Ahima RS, Saper CB, Flier JS, Elmquist JK. Leptin regulation of neuroendocrine systems. *Front Neuroendocrinol.* (2000) 21:263–307. doi: 10.1006/frne.2000.0197
- Roa J, Tena-Sempere M. Energy balance and puberty onset: emerging role of central mTOR signaling. *Trends Endocrinol Metab.* (2010) 21:519–28. doi: 10.1016/j.tem.2010.05.003
- Feng Z, Zhang C, Kang HJ, Sun Y, Wang H, Naqvi A, et al. Regulation of female reproduction by p53 and its family members. *FASEB J.* (2011) 25:2245–55. doi: 10.1096/fj.10-180166
- Ko EK, Chorich LP, Sullivan ME, Cameron RS, Layman LC. JAK/STAT signaling pathway gene expression is reduced following Nelf knockdown in GnRH neurons. *Mol Cell Endocrinol.* (2018) 470:151–9. doi: 10.1016/j.mce.2017.10.009
- Stamatides GA, Kaiser UB. Gonadotropin regulation by pulsatile GnRH: signaling and gene expression. *Mol Cell Endocrinol.* (2018) 463:131–41. doi: 10.1016/j.mce.2017.10.015
- Manfredi-Lozano M, Roa J, Tena-Sempere M. Connecting metabolism and gonadal function: novel central neuroendocrine pathways involved in the metabolic control of puberty and fertility. *Front Neuroendocrinol.* (2018) 48:37–49. doi: 10.1016/j.yfrne.2017.07.008
- Corder R, Walker CD, Gaillard RG, Aubert ML. Inhibition of sexual maturation in male rats by melatonin: evidence linking the mechanism of action to changes in the regulation of hypothalamic neuropeptide y. *J Neuroendocrinol.* (1992) 4:1–8. doi: 10.1111/j.1365-2826.1992.tb00337.x
- Pierroz DD, Gruaz NM, d'Alviès V, Aubert ML. Chronic administration of neuropeptide Y into the lateral ventricle starting at 30 days of life delays sexual

maturation in the female rat. *Neuroendocrinology*. (1995) 61:293–300. doi: 10.1159/000126851

51. Janjic MM, Stojilkovic SS, Bjelobaba I. Intrinsic and regulated gonadotropin-releasing hormone receptor gene transcription in mammalian pituitary gonadotrophs. *Front Endocrinol*. (2017) 8:221. doi: 10.3389/fendo.2017.00221

52. Kraus S, Naor Z, Seger R. Intracellular signaling pathways mediated by the gonadotropin-releasing hormone (GnRH) receptor. *Arch Med Res*. (2001) 32:499–509. doi: 10.1016/S0188-4409(01)00331-9

53. Naor Z. Signaling by G-protein-coupled receptor (GPCR): studies on the GnRH receptor. *Front Neuroendocrinol*. (2009) 30:10–29. doi: 10.1016/j.yfrne.2008.07.001



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Gan Shangquan,
Xinjiang Academy of Agricultural and
Reclamation Sciences (XAARS), China
Jianning He,
Qingdao Agricultural University, China

*CORRESPONDENCE

Yongdong Peng
✉ pyongdong123@126.com
Changfa Wang
✉ wangchangfa@lcu.edu.cn

RECEIVED 31 March 2024

ACCEPTED 30 May 2024

PUBLISHED 05 July 2024

CITATION

Peng Y, Zhu M, Gong Y and Wang C (2024)
Identification and functional prediction of
lncRNAs associated with intramuscular lipid
deposition in Guangling donkeys.
Front. Vet. Sci. 11:1410109.
doi: 10.3389/fvets.2024.1410109

COPYRIGHT

© 2024 Peng, Zhu, Gong 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.

Identification and functional prediction of lncRNAs associated with intramuscular lipid deposition in Guangling donkeys

Yongdong Peng*, Mingzhu Zhu, Yunyi Gong and
Changfa Wang*

Liaocheng Research Institute of Donkey High-Efficiency Breeding and Ecological Feeding,
Agricultural Science and Engineering School, Liaocheng University, Liaocheng, China

Many studies have shown that long non-coding RNAs (lncRNAs) play key regulatory roles in various biological processes. However, the importance and molecular regulatory mechanisms of lncRNAs in donkey intramuscular fat deposition remain to be further investigated. In this study, we used published transcriptomic data from the longissimus dorsi muscle of Guangling donkeys to identify lncRNAs and obtained 196 novel lncRNAs. Compared with the coding genes, the novel lncRNAs and the known lncRNAs exhibited some typical features, such as shorter transcript length and smaller exons. A total of 272 coding genes and 52 lncRNAs were differentially expressed between the longissimus dorsi muscles of the low-fat and high-fat groups. The differentially expressed genes were found to be involved in various biological processes related to lipid metabolism. The potential target genes of differentially expressed lncRNAs were predicted by cis and trans. Functional analysis of lncRNA targets showed that some lncRNAs may act on potential target genes involved in lipid metabolism processes and regulate lipid deposition in the longissimus dorsi muscle. This study provides valuable information for further investigation of the molecular mechanisms of lipid deposition traits in donkeys, which may improve meat traits and facilitate the selection process of donkeys in future breeding.

KEYWORDS

Guangling donkey, lncRNA, longissimus dorsi muscle, lipid deposition, regulate

1 Introduction

Long non-coding RNA (lncRNA) is a class of non-coding RNAs with a length greater than 200 nucleotides. Increasing evidence suggests that lncRNA plays important roles in various biological processes, such as embryonic development (1, 2), gene expression regulation (3, 4), reprogramming (5, 6), and genomic imprinting (7, 8). Additionally, many lncRNAs have been found to be involved in regulating lipid metabolism. For example, lncRNA IMFNCR promotes chicken myoblast differentiation by sequestering miR-128-3p and miR-27b-3p (9), lncRNA 332,443 inhibits preadipocyte differentiation by targeting Runx1, p38 MAPK, and ERK1/2-MAPK signaling pathways (10), and lncLSTR forms a molecular complex with TDP-43 to regulate the expression level of Cyp8b1, thereby affecting the FXR regulatory pathway, leading to increased apoC2 levels

and influencing triglyceride levels. LNCLSR directly binds to TDP-43 to inhibit Cyp8b1 expression and subsequently regulate triglyceride levels (11).

Guangling donkeys are distributed in Guangling County, Shanxi Province, China, and they are a local dominant breed that is carefully reared by local people using traditional production practices (12). Guangling donkeys have a stout physique and full muscles. Guangling donkeys used for meat production have a high intramuscular fat (IMF) content; however, the underlying molecular mechanisms underlying the IMF variation among donkey species are not fully understood.

IMF, also known as marbling, is an important indicator of the lean meat-to-fat ratio, which directly affects the tenderness, juiciness, and flavor of the meat (13). IMF content is one of the most important indicators used to evaluate meat quality (14). IMF content is a polygenic trait that is regulated by many genes affecting adipogenesis and lipid metabolism (15). At present, the underlying molecular variations affecting IMF content among donkey breeds are unclear.

This study used published transcriptome data from the longissimus dorsi muscle of a Guangling donkey to identify lncRNA and conducted differential expression studies of coding genes and lncRNA, constructing an expression regulation network, which lays the foundation for further analyzing the molecular mechanisms of lipid deposition traits in donkeys.

2 Materials and methods

2.1 Data sources

A total of 30 Guangling donkeys were raised on a commercial donkey farm in Fanshi County, Xinzhou City, Shanxi Province, China, and 6 donkeys with IMF differences and similar ages were selected (age: 2–3 years old, weight: 232–245 kg; female) for use in this study. All Guangling donkeys were reared under the same natural conditions of uncontrolled room temperature and light with unrestricted access to food and water. The longissimus dorsi samples at the 13th rib were aseptically and quickly obtained within 30 min of harvest. The collected samples were stored in liquid nitrogen for immediate storage, and long-term storage was carried out at -80°C . According to the China National Standard GB5009.6-2016 “Determination of Fat in Foods in National Food Safety Standard,” the IMF content was determined by the Soxhlet method. A Soxhlet extraction apparatus was used to remove fat and dry the ground meat samples for fat extraction. Petroleum ether was used as a solvent. This was recycled and dried for 8 h. Then, it was weighed to obtain the weight of the bottle containing fat. The IMF content was calculated by a formula. The three longissimus dorsi samples with the highest IMF contents and the other three with the lowest IMF contents were selected for transcriptome analysis (12). A total of six RNA-seq datasets were obtained from a previously published study and downloaded from the NCBI’s GEO database (PRJNA658642) (12). The donkey gene annotations were downloaded from https://ftp.ensembl.org/pub/release-110/gtf/equus_asinus. Moreover, the Non-Redundant Protein Sequence (NR) Database was downloaded from <ftp://ftp.ncbi.nih.gov/blast/db/>. The uniref90 database was

downloaded from <https://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref90>.

2.2 RNA-seq reads mapping and transcriptome assembly

The quality of sequencing reads was evaluated by FastQC command. The raw reads were filtered and trimmed by Trimmomatic (version 0.39) with default parameters (16). The clean reads were then mapped to the donkey reference genome [Ensembl: ASM1607732v2 (GCA_016077325.2)] by HISAT2 v2.2.1 with the default parameters (17–20). StringTie (version 2.2.1) was used to assemble the mapped reads with default parameters (17). Then, the merge tool of StringTie was used to merge the six assembled transcript files (GTF format) of the two groups into a non-redundant transcriptome. In addition, by using the assembled GTF file, StringTie software was used to estimate the expression levels of genes and transcripts in all samples for subsequent studies with the parameters “-e” and “-B” (17, 18, 21).

2.3 LincRNAs identification pipeline

The pipeline for lincRNA (long intergenic non-coding RNA) identification was as follows (Figure 1): (1) retained those transcripts with “u” category categorized by using gffcompare, which indicated intergenic transcripts (20, 21). (2) According to the merged GTF file, the transcripts with single exons and less than 200 bp in length were removed (18, 21). (3) The CPC2, CNCI, PLEK and LGC were used to assess the protein-coding potential of complete transcript sequences, and the transcripts that cannot encode proteins based on protein-coding potential were retained (22). (4) The HMMER was used to identify the transcripts translated in all six possible frames with homologs that were concluded in any of the known protein family domains in the Pfam database, and transcripts that matched to the Pfam hit ($E\text{-value} < 1e-5$) were excluded (18–21, 23). (5) BLASTX program (24) was used to filter out any transcripts that have similarities to known proteins in the NCBI NR and UniRef90 databases ($E\text{-value} < 1e-5$) (20, 21). (6) Reserve transcripts with FPKM values greater than 0 in at least one sample (18–20).

2.4 Comparisons between lncRNAs and protein-coding transcripts

We selected the transcripts annotated as “protein-coding” in the gene annotation file, and the obtained lncRNAs were screened with “known” and “novel” by “blastn” command. The transcript length, exon length, and exon number of lncRNAs were compared with those of protein-coding transcripts (20).

2.5 Analysis of differentially expressed genes and differentially expressed lncRNAs

DESeq2 tool was used to perform differential expression analysis of protein-coding genes and lncRNAs between the high (H) fat group and low (L) fat group (25). $|\log_2 \text{fold change}| \geq 1$ and adjusted $p\text{-value}$

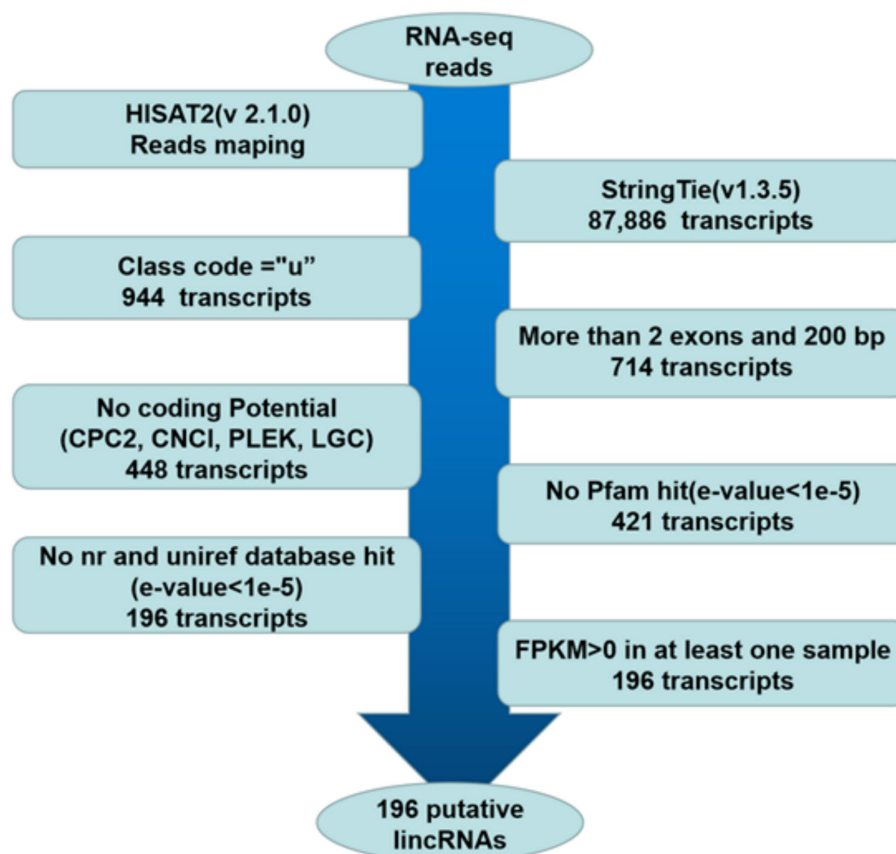


FIGURE 1

Pipeline for the identification of putative lincRNAs in this study. The frames in the direction of the arrow show the filtering process and the number of screened transcripts. "u": Unknown, intergenic transcript; CPC2: coding potential calculator 2; nr: non-redundant protein sequence database; FPKM: fragments per kilobase of transcript per million mapped reads.

(padj) < 0.05 were used to screen differentially expressed genes (DEGs) and lincRNAs (26).

2.6 Prediction of potential target genes

We predicted the molecular functions of protein-coding genes regulated by RNA in cis and trans. First, the neighboring protein-coding genes near DELs (<100kb) were identified based on cis-prediction principles using Bedtools (18–20, 27, 28). For the trans-regulation of DELs, we calculated the Pearson's correlation coefficient (r) between DELs and protein-coding genes. We selected protein-coding genes with a Pearson's correlation coefficient $|r| \geq 0.95$, p -value ≤ 0.01 as potential target genes (PTGs) of DELs (21, 29).

2.7 Functional enrichment analysis

Gene Ontology (GO) enrichment analysis was performed by clusterProfiler (30, 31). KOBAS v3.0¹ was used for the Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (32). A p -value of less than 0.05 was considered statistically significant (33).

3 Results

3.1 Summary of RNA-seq data mapping and transcripts assembly in longissimus dorsi muscles

RNA-seq data involving two groups of Guangling donkeys were obtained from a previously published study (12). The six longissimus dorsi muscles (three in each group) with the lowest and highest IMF contents were named the L group and H group, respectively. The individual samples in the groups were named L1, L2, L3, H1, H2, and H3. The clean reads were mapped to the donkey reference genome using HISAT2. Approximately 86.02–90.51% of clean reads from each library were mapped to the donkey reference genome, and 79.76–84.78% of the reads were uniquely mapped to the genome. Then, the transcriptome was assembled for each library by StringTie, and all transcripts were synthesized into non-redundant transcripts using StringTie-Merge. After merging non-redundant transcripts, approximately 1.07% (944 of 87,886) of the transcripts were intergenic

¹ <http://kobas.cbi.pku.edu.cn>

transcripts. The 196 putative lincRNAs were obtained according to the illustration shown in [Figure 1](#) ([Supplementary Table S1](#)).

3.2 Comparison of coding genes and lncRNA features

Previous studies showed that there are many differences between protein-coding transcripts and lncRNAs ([18–20](#), [27](#), [34–36](#)). According to the assembled transcriptome, the characteristics of lncRNA and protein-coding transcripts were compared. A total of 51,112 protein-coding transcripts, corresponding to 20,553 protein-coding genes annotated in donkeys, were acquired. In addition, the donkey annotation file contains 7,615 known lncRNA transcripts that correspond to 4,709 lncRNA genes ([Supplementary Table S1](#)).

The average transcript length of the protein-coding transcripts (2,919 bp) was longer than the novel lncRNA transcripts (1,217 bp) and the known lncRNA transcripts (2,394 bp) ([Figure 2A](#)). In terms of average exon length, the novel lncRNA is 448 bp in length, shorter than the known lncRNA gene (766 bp) but longer than the protein-coding transcripts (271 bp) ([Figure 2B](#)). In addition, we found that the average number of exons in the protein-coding transcripts is 10.7, which is significantly higher than that of the novel lncRNA (2.7) and the known lncRNA (3.1) ([Figure 2C](#)).

3.3 Differential expression analysis of coding genes and lncRNAs

To explore their potential biological functions, we performed differential expression analysis of coding genes and lncRNAs. A total of 272 DEGs were obtained by comparing the high-fat group with the low-fat group of the longissimus dorsi muscle samples. Among them, 147 genes were upregulated in the high-fat group, and 125 genes were downregulated ([Figure 3A](#)). Additionally, 52 differentially expressed lncRNAs (DELs) were identified, with 25 lncRNAs upregulated and 27 lncRNAs downregulated in the high-fat group compared to the low-fat group ([Figure 3B](#)).

3.4 Functional analysis of differentially expressed genes

To investigate the function of DEGs, we performed GO and KEGG analyses, respectively. GO analysis of DEGs showed that integrin-mediated signaling pathway, cellular response to lipopolysaccharide, trachea development, cellular response to molecule of bacterial origin, and positive regulation of myeloid leukocyte-mediated immunity were the most abundant terms in the biological process category. In terms of cellular component category, actin cytoskeleton, phagocytic cup, cell leading edge, TORC2 complex, and basolateral plasma membrane were the top five terms, while rRNA binding, pattern recognition receptor activity, motor activity, protein kinase binding, and ion channel binding were most prevalent in the molecular function ([Supplementary Table S2](#)). Some GO terms were significantly associated with lipid metabolisms, such as response to lipids, cellular response to lipids, glycerophospholipid metabolic process, glycerolipid metabolic process, unsaturated fatty acid biosynthetic process, and phospholipid dephosphorylation. KEGG

analysis indicated that DEGs were significantly enriched in 62 KEGG pathways, of which several pathways were related to lipid metabolism, such as the Sphingolipid signaling pathway and MAPK signaling pathway. In addition, some other pathways are related to lipid metabolism, namely fatty acid metabolism, glycerophospholipid metabolism, biosynthesis of unsaturated fatty acids, PI3K-Akt signaling pathway, fatty acid degradation, ether lipid metabolism, cholesterol metabolism, PPAR signaling pathway, TGF-beta signaling pathway, and Wnt signaling pathway ([Supplementary Table S3](#)).

3.5 Prediction and functional analysis of lncRNA target genes

Many studies have indicated that lncRNA may regulate adjacent genes in a *cis* manner ([18–20](#), [29](#), [37–41](#)). For PTGs regulated by lncRNAs in *cis* (<100 kb), we identified a total of 323 PTGs, corresponding to 52 DELs ([Supplementary Table S4](#)). To explore the function of putative lncRNAs, GO and KEGG analyses were performed on expressed protein-coding genes transcribed near lncRNA (<100 kb). The results indicated that 102 of 323 PTGs were significantly involved in 143 biological processes ([Supplementary Table S5](#)). Of them, 59 PTGs significantly participated in 32 pathways ([Supplementary Table S6](#)), including 2 pathways related to lipid metabolism, such as biosynthesis of unsaturated fatty acids and fatty acid metabolism.

lncRNAs can not only regulate the expression of neighboring protein-coding genes through a *cis* mechanism but also regulate the expression of genes located on other chromosomes via a *trans* mechanism ([42](#)). In this study, we carried out the *trans* analysis to find the PTGs that were significantly correlated ($|r| \geq 0.95$, $p \leq 0.01$) to the DELs. In total, 3,366 PTGs were highly correlated with 52 DELs. Among these genes, 132 PTGs were differentially expressed in groups as DEPTGs, suggesting that most of the lncRNAs regulated gene expression through *trans* regulation. GO enrichment analysis showed that 3,366 PTGs were enriched in 835 biological processes and 132 DEPTGs were enriched in 143 biological processes ([Supplementary Tables S7, S8](#)). In cases of biological process, some GO terms were significantly associated with lipid metabolism, such as negative regulation of the lipid catabolic process, regulation of the lipid metabolic process, negative regulation of the lipid metabolic process, regulation of the lipid biosynthetic process, cellular lipid catabolic process, negative regulation of the lipid biosynthetic process, regulation of fatty acid transport and regulation of the lipid catabolic process ([Figures 4A,C](#)). In addition, 3,366 PTGs and 132 DEPTGs were enriched in 74 pathways and 7 pathways, respectively ([Supplementary Tables S9, S10](#)). KEGG pathways were involved in the MAPK signaling pathway, PI3K-Akt signaling pathway, glycerolipid metabolism, ether lipid metabolism, fat digestion, and absorption ([Figures 4B,D](#)). The results indicated that DELs had an important role in regulating their PTGs that regulate lipid metabolism in longissimus dorsi muscles.

4 Discussion

In the present study, transcriptome sequencing of longissimus dorsi tissues with different IMF contents from Guangling donkeys was used to investigate genes and lncRNA related to lipid metabolism in the longissimus dorsi muscle. A total of 20,553 protein-coding genes,

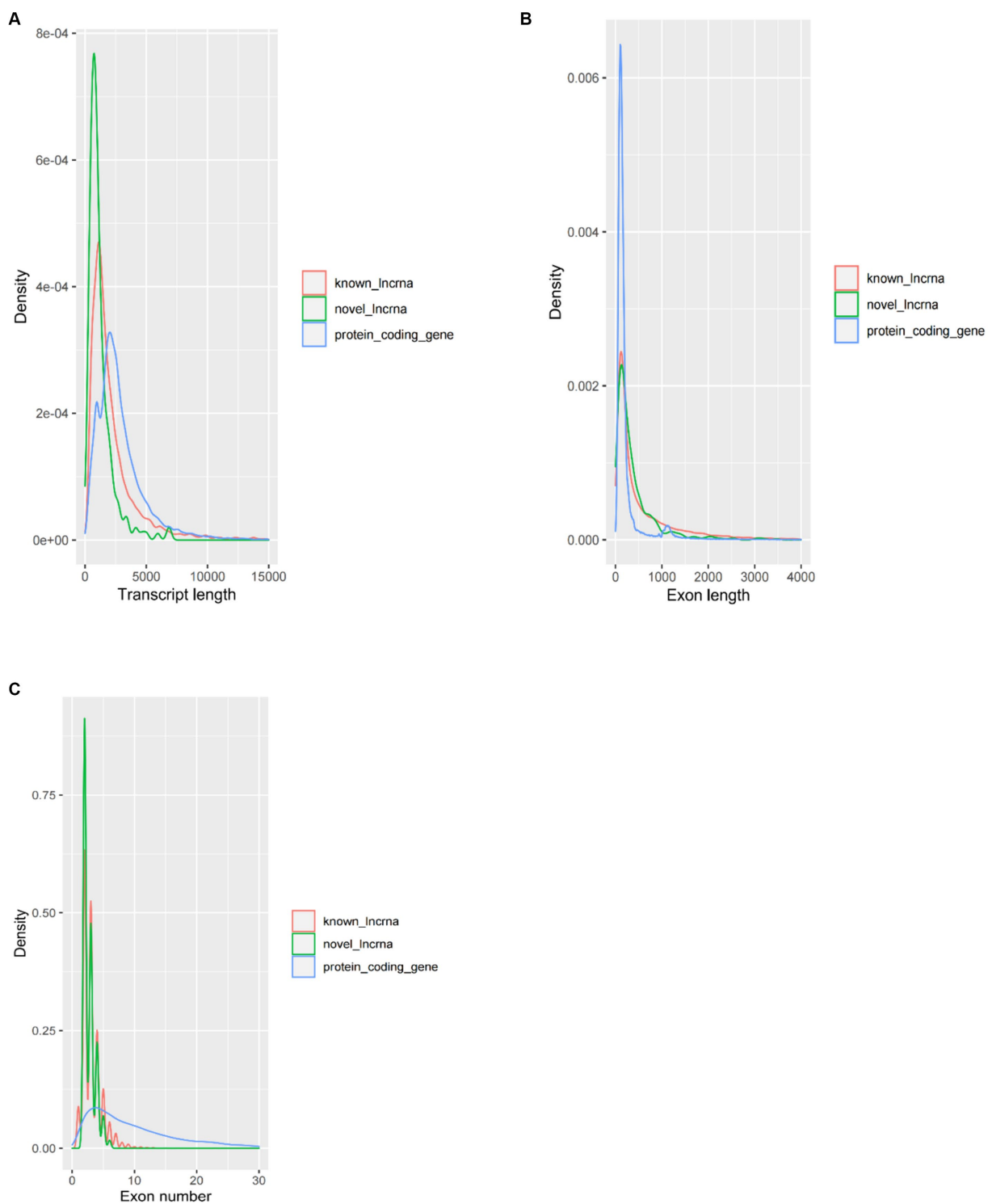


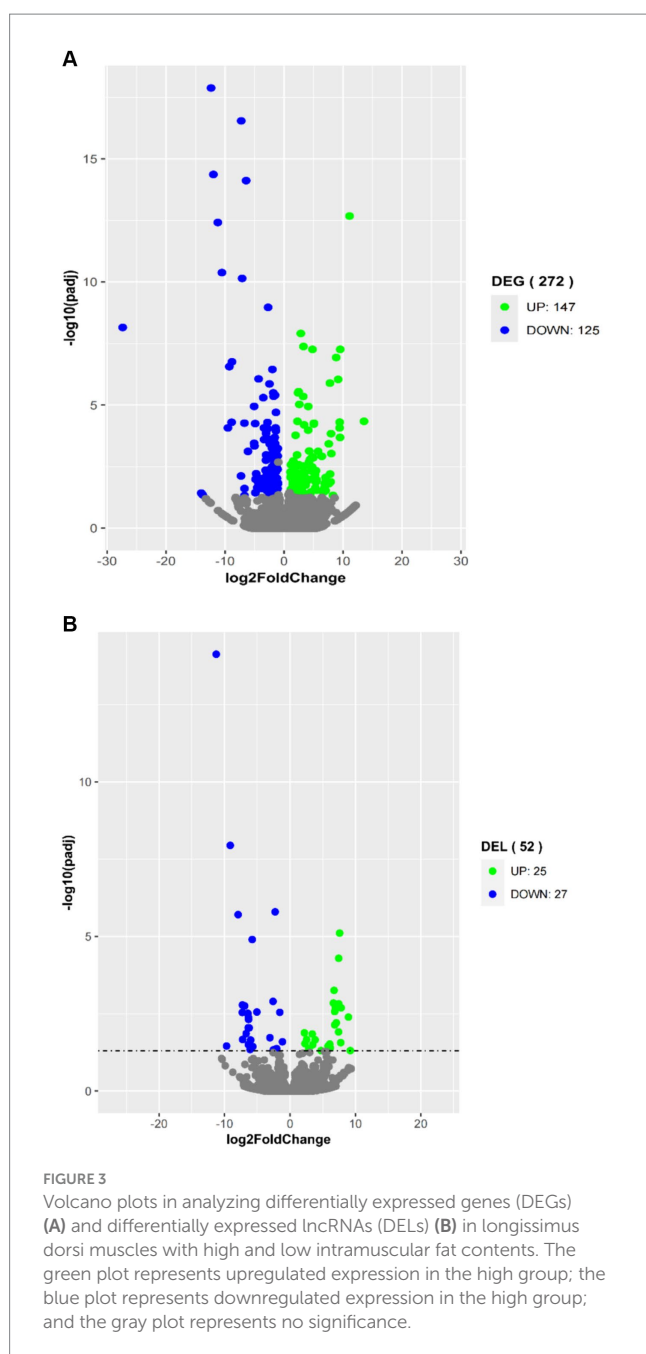
FIGURE 2

Comparison of the characteristics of protein-coding genes and lncRNA genes. (A) Comparison of transcript length, (B) comparison of exon length, and (C) comparison of exon number.

7,615 known lncRNAs, and 196 novel lincRNAs were obtained. We compared the known and novel lincRNAs with the donkey protein-coding genes and found that known and novel lincRNAs have shorter transcript lengths, longer exon lengths, and fewer exon

numbers compared with protein-coding transcripts, which is consistent with some previous studies (18–20, 27, 34–36, 43–46).

Differential expression analysis identified 272 DEGs, with 147 genes upregulated and 125 genes downregulated in the high-fat group



compared to the low-fat group. Some of these genes may play a key role in lipid metabolism. As expected, GO analysis revealed the involvement of a significant number of DEGs in lipid metabolism-related biological processes, including response to lipid, cellular response to lipid, glycerophospholipid metabolic process, glycerolipid metabolic process, unsaturated fatty acid biosynthetic process, and phospholipid dephosphorylation. KEGG analysis showed significant enrichment of 62 KEGG pathways in the DEGs. Enriched pathways included the sphingolipid signaling pathway, MAPK signaling pathway, fatty acid metabolism, glycerophospholipid metabolism, biosynthesis of unsaturated fatty acids, PI3K-Akt signaling pathway, fatty acid degradation, ether lipid metabolism, cholesterol metabolism, PPAR signaling pathway, TGF-beta signaling pathway, and Wnt signaling pathway, which are involved in lipid metabolism.

Previous studies indicated that lncRNAs can regulate gene expression in a cis-acting manner (18, 20, 29, 35, 41). In the present study, a total of 196 novel lncRNAs and 7,615 known lncRNAs, were identified. Moreover, 52 DELs were detected, 25 of which were upregulated and 27 downregulated in the high-fat group compared with the low-fat group. To predict the function of these lncRNAs, protein-coding genes transcribed near lncRNAs (<100 kb) were screened. A total of 323 PTGs were identified, corresponding to 52 DELs. Exploration of lncRNA function through GO and KEGG analysis of PTGs revealed that 102 of 323 PTGs were significantly involved in 143 biological processes, and 59 PTGs significantly participated in 7 pathways, including 2 pathways related to lipid metabolism, such as biosynthesis of unsaturated fatty acids and fatty acid metabolism. In addition, four other pathways are related to lipid metabolism, namely the Wnt signaling pathway, glycerolipid metabolism, PPAR signaling pathway, and glycerophospholipid metabolism. These pathways can affect lipid metabolism to some extent.

An intriguing observation is that the same DEL ENSEAST00005078264 acts on both FADS1 and FADS2, which are the key enzymes that catalyze adenylation of flavin mononucleotide (FMN) to form flavin adenine dinucleotide (FAD) coenzyme (47, 48). These two PTGs are involved in the biosynthesis of unsaturated fatty acids and fatty acid metabolism. The DEL ENSEAST00005078264 may be involved in lipid metabolism via biosynthesis of unsaturated fatty acids and fatty acid metabolism signaling pathway. Another DEL ENSEAST00005069204 was found to target AGPAT5, which converts lysophosphatidic acid to phosphatidic acid, the second step in *de novo* phospholipid biosynthesis (49). These lncRNAs may regulate target genes involved in lipid metabolism pathways through cis-acting mechanisms.

The analysis of PTGs by trans-acting lncRNAs was conducted, and enriched analysis was also performed. Among the 3,366 PTGs, 1,352 were significantly involved in 835 biological processes. Pathway analysis showed that 710 PTGs were significantly involved in 74 pathways, including MAPK signaling pathway, PI3K-Akt signaling pathway, glycerolipid metabolism, ether lipid metabolism, fat digestion, and absorption, all of which are related to lipid metabolism. Other pathways related to lipid metabolisms, such as glycerophospholipid metabolism, fatty acid metabolism, biosynthesis of unsaturated fatty acids, sphingolipid signaling pathway, cholesterol metabolism, adipocytokine signaling pathway, Wnt signaling pathway, fatty acid elongation, regulation of lipolysis in adipocytes, sphingolipid metabolism, TGF-beta signaling pathway, fatty acid biosynthesis, and fatty acid degradation, were also identified. For 132 DEPTG, 56 were significantly involved in 143 biological processes. Pathway analysis showed that 27 DEPTGs were significantly involved in 7 pathways, including arachidonic acid metabolism. Other pathways involved in lipid metabolism have also been identified, including biosynthesis of unsaturated fatty acids, fat digestion and absorption, ether lipid metabolism, fatty acid metabolism, MAPK signaling pathway, PPAR signaling pathway, glycerophospholipid metabolism, and PI3K-Akt signaling pathway.

It was found that SCD, targeted by four DELs (ENSEAST00005042127, ENSEAST00005051768, ENSEAST00005052324, and ENSEAST00005072263), is a key gene that regulates lipid metabolism. Studies have shown that SCD plays an important role in lipid biosynthesis (50–52). In addition, two DELs,

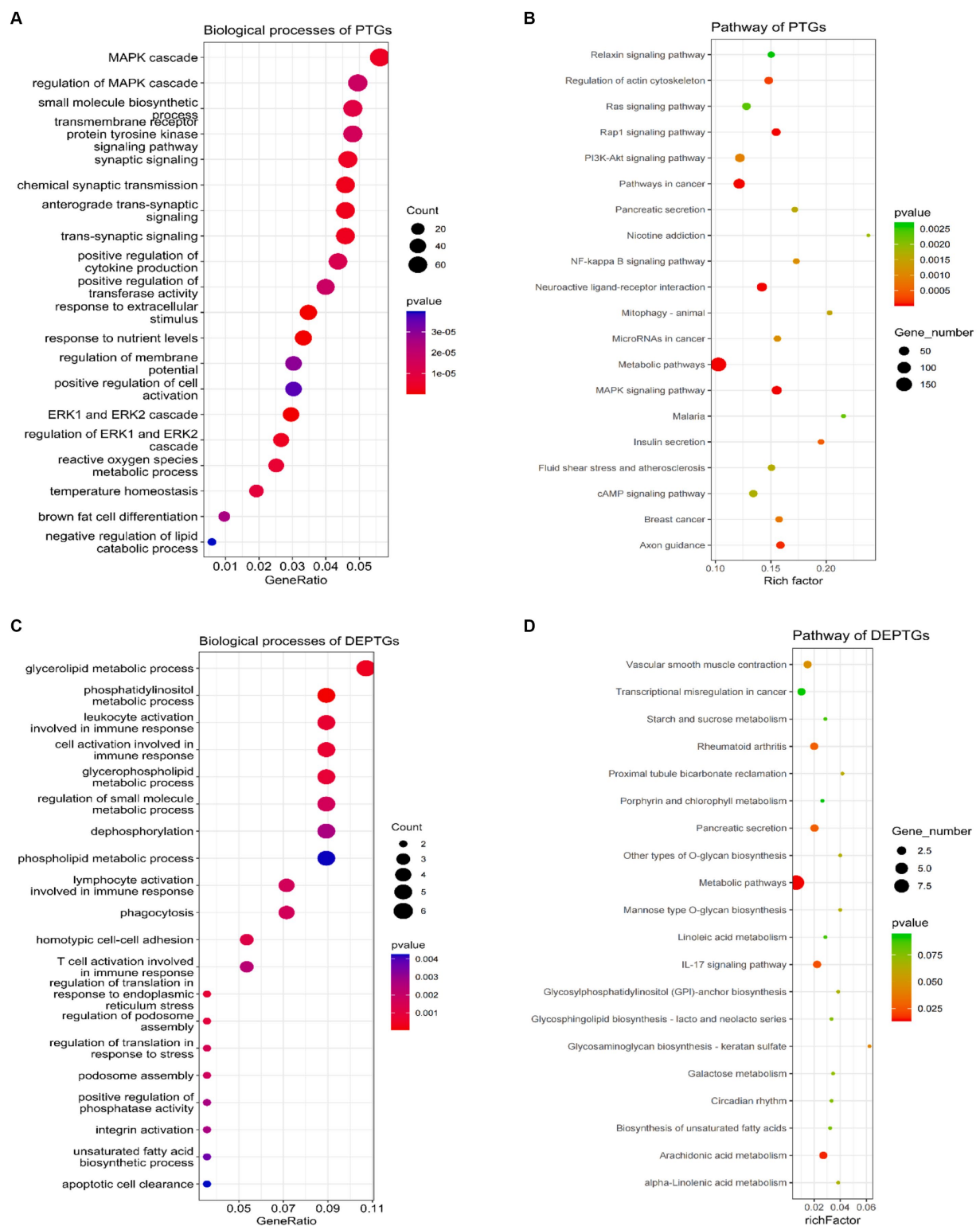


FIGURE 4

Gene Ontology and pathway analysis of PTGs of DELs. (A) Biological processes analysis of PTGs of DELs, (B) pathway analysis of PTGs of DELs, (C) biological processes analysis of DEPTGs of DELs, and (D) pathway analysis of DEPTGs of DELs. PTGs, potential target genes; DELs, differentially expressed lncRNAs; DEPTGs, differentially expressed potential target genes.

ENSEAST00005052438 and ENSEAST00005076477, were all found to act on PLA2G3, which is involved in lipid metabolism and catalyzes the calcium-dependent hydrolysis of the sn-2 acyl bond of

phospholipids to release arachidonic acid and lysophospholipids (53–56). It is noteworthy that THRSP was targeted by four DELs (ENSEAST00005042127, ENSEAST00005048792,

ENSEAST00005051768, and ENSEAST00005052324). Previous studies showed that THRSP is important for the biosynthesis of triglycerides with medium-length fatty acid chains and plays a role in the regulation of lipogenesis, which may modulate lipogenesis by interacting with MID1IP1 and preventing its interaction with ACACA, may function as a transcriptional coactivator and may modulate the transcription factor activity of THRSP (57, 58). These lncRNAs may regulate target genes involved in lipid metabolism pathways through trans-acting mechanisms.

SCD plays an important role in the regulation of lipid deposition, and it can catalyze the conversion of saturated fatty acids (SFAs) to monounsaturated fatty acids (MUFAs) (59). It was found that SCD expression was positively correlated with SCD protein levels, and pigs with higher SCD protein levels had higher levels of IMF, indicating a conserved trend between transcriptional and translational levels (60). Smith et al. (61) found that SCD expression was closely related to marbled adipocyte differentiation and that grain diets increased SCD expression, resulting in higher intracellular levels of MUFA (especially oleic acid), and thus increased adipocyte differentiation. In addition, increased FA levels in IMF tissues were found to be associated with tissue-specific activation of SCD expression under the influence of a low-protein protein diet, and the low-protein protein diet significantly increased the expression and activity of SCD proteins in muscle, but not in subcutaneous adipose tissue (AT) (62). These results suggest that the high specific expression of SCD in IMF cells may be related to meat quality and that reduced protein intake may affect IMF content by modulating SCD activity. Notably, SCD was significantly up-regulated in the high-fat group compared with the low-fat group in this study. The results show that SCD may play an important function in donkey IMF deposition. The thyroid hormone-sensitive protein (THRSP; Spot14; S14) is a nuclear protein that is abundantly expressed in lipogenic tissues such as in the liver, mammary gland, AT and lipogenic breast cancers (63–66). A previous study indicated that the polymorphisms and genotype distribution of THRSP were closely related to the potential for fat production in pig breeds (67). It was found that miR-195 may inhibit lipid accumulation in adipocytes by regulating THRSP (68). THRSP is regulated by insulin both *in vivo* in human AT and *in vitro* in adipocytes, and its expression is downregulated by insulin resistance. As THRSP silencing decreases mitochondrial respiration and fatty acid oxidation, its downregulation in human AT could contribute to mitochondrial dysfunction. Furthermore, disturbed sphingolipid metabolism could contribute to metabolic dysfunction in obese AT (69). The current RNA-seq results revealed that THRSP had higher expression in the high-fat group than in the low-fat group, which suggested that THRSP may be a key gene involved in IMF deposition in donkeys. Taken together, some DEGs, including SCD and THRSP, may be key candidate genes for donkey meat quality improvement.

It is noteworthy that SCD was targeted by four DELs (ENSEAST00005042127, ENSEAST00005051768, ENSEAST00005052324, and ENSEAST00005072263). Moreover, four DELs (ENSEAST00005042127, ENSEAST00005048792, ENSEAST00005051768, and ENSEAST00005052324) were found to target THRSP. These lncRNAs may influence donkey IMF deposition by regulating the expression of their target genes. The relationship of these lncRNAs with their target genes and with fat deposition in donkey muscle needs to be further investigated in the future using molecular biology, gene editing, etc. These lncRNAs may improve

meat quality and facilitate the selection process of donkeys in future breeding.

5 Conclusion

In the study, we identified 196 putative lncRNAs and analyzed the characteristics of lncRNAs compared with protein-coding genes in the longissimus dorsi muscles of Guangling donkeys. We observed numerous DELs and protein-coding genes in longissimus dorsi muscles with different IMF contents. Some DEGs were found to be involved in various biological processes related to lipid metabolism. Functional enrichment analysis of PTGs by DELs revealed that some lncRNAs (such as ENSEAST00005042127, ENSEAST00005051768, ENSEAST00005052324, ENSEAST00005072263, ENSEAST00005052438, ENSEAST00005076477, ENSEAST00005042127, ENSEAST00005048792, ENSEAST00005051768, and ENSEAST00005052324) may act on PTGs (such as SCD, PLA2G3, and THRSP), participate in lipid metabolism processes, and regulate IMF deposition in the longissimus dorsi muscle. This study provides valuable resources for future analyses of lipid deposition traits and may contribute to the improvement of donkey meat quality and the selection process in donkey 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 studies were approved by Animal Ethics Committee of Liaocheng University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

YP: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. MZ: Data curation, Investigation, Methodology, Visualization, Writing – review & editing. YG: Data curation, Formal analysis, Methodology, Software, Writing – review & editing. CW: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Shandong Province Modern Agricultural Industrial

Technology System Project (grant number SDAIT-27), the Construction of Molecular ID Card for Donkey and Camel Species, the Livestock and Poultry Seed Industry Project of the Ministry of Agriculture and Rural Affairs (grant number 19211162), the Key R&D Project of Shandong Province: Innovation and Demonstration of Key Technologies for the Integrated Development of the Donge Black Donkey Industry (grant number 2021TZXD012), and the Liaocheng University Scientific Research Fund (grant number 318052339).

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.

References

- Zhang K, Huang K, Luo Y, Li S. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC Genomics*. (2014) 15:845. doi: 10.1186/1471-2164-15-845
- Li H, Wang Z, Zhao B, Zhang H, Fan D, Ma H, et al. Sperm-borne lncRNA loc100847420 improves development of early bovine embryos. *Anim Reprod Sci*. (2023) 257:107333. doi: 10.1016/j.anireprosci.2023.107333
- Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol*. (2018) 19:143–57. doi: 10.1038/nrm.2017.104
- Natarajan P, Shrinivas K, Chakraborty AK. A model for cis-regulation of transcriptional condensates and gene expression by proximal lncRNAs. *Biophys J*. (2023) 122:2757–72. doi: 10.1016/j.bpj.2023.05.032
- Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. (2010) 42:1113–7. doi: 10.1038/ng.710
- He XY, Fan X, Qu L, Wang X, Jiang L, Sang LJ, et al. lncRNA modulates hippo-yap signaling to reprogram iron metabolism. *Nat Commun*. (2023) 14:2253. doi: 10.1038/s41467-023-37871-5
- Prickett AR, Oakey RJ. A survey of tissue-specific genomic imprinting in mammals. *Mol Gen Genomics*. (2012) 287:621–30. doi: 10.1007/s00438-012-0708-6
- Di Michele F, Chillon I, Feil R. Imprinted long non-coding RNAs in mammalian development and disease. *Int J Mol Sci*. (2023) 24:13647. doi: 10.3390/ijms241713647
- Zhang M, Li F, Sun JW, Li DH, Li WT, Jiang RR, et al. lncRNA IMFNCR promotes intramuscular adipocyte differentiation by sponging miR-128-3p and miR-27b-3p. *Front Genet*. (2019) 10:42. doi: 10.3389/fgene.2019.00042
- Xiao F, Tang CY, Tang HN, Wu HX, Hu N, Li L, et al. Long non-coding RNA 332443 inhibits Preadipocyte differentiation by targeting Runx1 and P38-Mapk and Erk1/2-Mapk signaling pathways. *Front Cell Dev Biol*. (2021) 9:663959. doi: 10.3389/fcell.2021.663959
- Li P, Ruan X, Yang L, Kiesewetter K, Zhao Y, Luo H, et al. A liver-enriched Long non-coding RNA, lncLSTR, regulates systemic lipid metabolism in mice. *Cell Metab*. (2015) 21:455–67. doi: 10.1016/j.cmet.2015.02.004
- Li W, Qiu L, Guan J, Sun Y, Zhao J, Du M. Comparative transcriptome analysis of longissimus Dorsi tissues with different intramuscular fat contents from Guangling donkeys. *BMC Genomics*. (2022) 23:644. doi: 10.1186/s12864-022-08857-2
- van Laack RL, Stevens SG, Stalder KJ. The influence of ultimate PH and intramuscular fat content on pork tenderness and tenderization. *J Anim Sci*. (2001) 79:392–7. doi: 10.2527/2001.792392x
- Jeremiah LE, Dugan ME, Aalhus JL, Gibson LL. Assessment of the chemical and cooking properties of the major beef muscles and muscle groups. *Meat Sci*. (2003) 65:985–92. doi: 10.1016/S0309-1740(02)00308-X
- Michal JJ, Zhang ZW, Gaskins CT, Jiang Z. The bovine fatty acid binding protein 4 gene is significantly associated with marbling and subcutaneous fat depth in Wagyu X Limousin F2 crosses. *Anim Genet*. (2006) 37:400–2. doi: 10.1111/j.1365-2052.2006.01464.x
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. (2014) 30:2114–20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170
- 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
- Chen L, Shi G, Chen G, Li J, Li M, Zou C, et al. Transcriptome analysis suggests the roles of Long intergenic non-coding RNAs in the growth performance of weaned piglets. *Front Genet*. (2019) 10:196. doi: 10.3389/fgene.2019.00196
- Chen G, Cheng X, Shi G, Zou C, Chen L, Li J, et al. Transcriptome analysis reveals the effect of Long intergenic noncoding RNAs on pig muscle growth and fat deposition. *Biomed Res Int*. (2019) 2019:2951427. doi: 10.1155/2019/2951427
- Shi G, Chen L, Chen G, Zou C, Li J, Li M, et al. Identification and functional prediction of Long intergenic non-coding RNAs related to subcutaneous adipose development in pigs. *Front Genet*. (2019) 10:160. doi: 10.3389/fgene.2019.00160
- Zhao W, Li Z, Liu Q, Xie S, Li M, Wang Y, et al. Analysis of Long intergenic non-coding RNAs transcriptomic profiling in skeletal muscle growth during porcine embryonic development. *Sci Rep*. (2021) 11:15240. doi: 10.1038/s41598-021-94014-w
- Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. Cpc2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*. (2017) 45:W12–6. doi: 10.1093/nar/gkx428
- Prakash A, Jeffreys M, Bateman A, Finn RD. The Hmmer web server for protein sequence similarity search. *Curr Protoc Bioinformatics*. (2017) 60:3 15 1–3 23. doi: 10.1002/cpbi.40
- Pirooznia M, Perkins EJ, Deng Y. Batch blast extractor: an automated Blastx parser application. *BMC Genomics*. (2008) 9:S10. doi: 10.1186/1471-2164-9-S2-S10
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with Deseq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8
- Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. (2001) 125:279–84. doi: 10.1016/S0166-4328(01)00297-2
- Zou C, Li L, Cheng X, Li C, Fu Y, Fang C, et al. Identification and functional analysis of Long intergenic non-coding RNAs underlying intramuscular fat content in pigs. *Front Genet*. (2018) 9:102. doi: 10.3389/fgene.2018.00102
- Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. (2011) 25:1915–27. doi: 10.1101/gad.17446611
- Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting Omics data. *Innovation (Camb)*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141
- Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. (2011) 39:W316–22. doi: 10.1093/nar/gkr483
- Benjamini Y, Cohen R. Weighted false discovery rate controlling procedures for clinical trials. *Biostatistics*. (2017) 18:91–104. doi: 10.1093/biostatistics/kxw030
- Zou C, Li J, Luo W, Li L, Hu A, Fu Y, et al. Transcriptome analysis reveals Long intergenic non-coding RNAs involved in skeletal muscle growth and development in pig. *Sci Rep*. (2017) 7:8704. doi: 10.1038/s41598-017-07998-9
- Zou C, Li S, Deng L, Guan Y, Chen D, Yuan X, et al. Transcriptome analysis reveals Long intergenic noncoding RNAs contributed to growth and meat quality differences

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.2024.1410109/full#supplementary-material>

between Yorkshire and Wannanhuo pig. *Genes (Basel)*. (2017) 8:203. doi: 10.3390/genes8080203

36. Gong Y, He J, Li B, Xiao Y, Zeng Q, Xu K, et al. Integrated analysis of lncRNA and mRNA in subcutaneous adipose tissue of Ningxiang pig. *Biology (Basel)*. (2021) 10:726. doi: 10.3390/biology10080726

37. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet*. (2020) 21:102–17. doi: 10.1038/s41576-019-0184-5

38. Zhou H, Simion V, Pierce JB, Haemmig S, Chen AF, Feinberg MW. LncRNA-Map3k4 regulates vascular inflammation through the P38 Mapk Signaling pathway and cis-modulation of Map3k4. *FASEB J*. (2021) 35:e21133. doi: 10.1096/fj.202001654RR

39. Szafranski P, Gambin T, Karolak JA, Popek E, Stankiewicz P. Lung-specific distant enhancer cis regulates expression of Foxf1 and lncRNA FENDRR. *Hum Mutat*. (2021) 42:694–8. doi: 10.1002/humu.24198

40. Degani N, Lubelsky Y, Perry RB, Ainbinder E, Ulitsky I. Highly conserved and cis-acting lncRNAs produced from paralogous regions in the Center of Hoxa and Hoxb Clusters in the endoderm lineage. *PLoS Genet*. (2021) 17:e1009681. doi: 10.1371/journal.pgen.1009681

41. Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G. Genome-wide identification, characterization and evolutionary analysis of Long intergenic noncoding RNAs in cucumber. *PLoS One*. (2015) 10:e0121800. doi: 10.1371/journal.pone.0121800

42. Feng H, Liu T, Yousuf S, Zhang X, Huang W, Li A, et al. Identification and analysis of lncRNA, miRNA and mRNA related to subcutaneous and intramuscular fat in Laiwu pigs. *Front Endocrinol*. (2022) 13:1081460. doi: 10.3389/fendo.2022.1081460

43. Yu H, Zhao X, Li Q. Genome-wide identification and characterization of long intergenic noncoding RNAs and their potential association with larval development in the Pacific oyster. *Sci Rep*. (2016) 6:20796. doi: 10.1038/srep20796

44. Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, et al. Systematic identification and characterization of long non-coding RNAs in the silkworm, *Bombyx Mori*. *PLoS One*. (2016) 11:e0147147. doi: 10.1371/journal.pone.0147147

45. Wang J, Fu L, Koganti PP, Wang L, Hand JM, Ma H, et al. Identification and prediction of large intergenic noncoding RNAs (linCRNAs) in rainbow trout (*Oncorhynchus Mykiss*). *Mar Biotechnol (NY)*. (2016) 18:271–82. doi: 10.1007/s10126-016-9689-5

46. Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, et al. Genome-wide screening and functional analysis identify a large number of Long noncoding RNAs involved in the sexual reproduction of Rice. *Genome Biol*. (2014) 15:512. doi: 10.1186/s13059-014-0512-1

47. Brizio C, Galluccio M, Wait R, Torchetti EM, Bafunno V, Accardi R, et al. Overexpression in *Escherichia coli* and characterization of two recombinant isoforms of human fad synthetase. *Biochem Biophys Res Commun*. (2006) 344:1008–16. doi: 10.1016/j.bbrc.2006.04.003

48. Olsen RKJ, Konarikova E, Giancaspero TA, Mosegaard S, Boczonadi V, Matakovic L, et al. Riboflavin-responsive and -non-responsive mutations in fad synthase cause multiple acyl-CoA dehydrogenase and combined respiratory-chain deficiency. *Am J Hum Genet*. (2016) 98:1130–45. doi: 10.1016/j.ajhg.2016.04.006

49. Prasad SS, Garg A, Agarwal AK. Enzymatic activities of the human Acpat isoform 3 and isoform 5: localization of Acpat5 to mitochondria. *J Lipid Res*. (2011) 52:451–62. doi: 10.1194/jlr.M007575

50. Wang J, Yu L, Schmidt RE, Su C, Huang X, Gould K, et al. Characterization of HSCD5, a novel human Stearoyl-CoA desaturase unique to primates. *Biochem Biophys Res Commun*. (2005) 332:735–42. doi: 10.1016/j.bbrc.2005.05.013

51. Goren MA, Fox BG. Wheat germ cell-free translation, purification, and assembly of a functional human Stearoyl-CoA desaturase complex. *Protein Expr Purif*. (2008) 62:171–8. doi: 10.1016/j.pep.2008.08.002

52. Zhang S, Yang Y, Shi Y. Characterization of human SCD2, an oligomeric desaturase with improved stability and enzyme activity by cross-linking in intact cells. *Biochem J*. (2005) 388:135–42. doi: 10.1042/BJ20041554

53. Murakami M, Masuda S, Shimbara S, Bezzine S, Lazdunski M, Lambeau G, et al. Cellular Arachidonate-releasing function of novel classes of secretory phospholipase A2s (groups iii and xii). *J Biol Chem*. (2003) 278:10657–67. doi: 10.1074/jbc.M211325200

54. Sato H, Kato R, Isogai Y, Saka G, Ohtsuki M, Taketomi Y, et al. Analyses of group iii secreted phospholipase A2 transgenic mice reveal potential participation of this enzyme in plasma lipoprotein modification, macrophage foam cell formation, and atherosclerosis. *J Biol Chem*. (2008) 283:33483–97. doi: 10.1074/jbc.M804628200

55. Murakami M, Masuda S, Shimbara S, Ishikawa Y, Ishii T, Kudo I. Cellular distribution, post-translational modification, and tumorigenic potential of human group iii secreted phospholipase a(2). *J Biol Chem*. (2005) 280:24987–98. doi: 10.1074/jbc.M502088200

56. Murase R, Taketomi Y, Miki Y, Nishito Y, Saito M, Fukami K, et al. Group iii phospholipase a(2) promotes colitis and colorectal cancer. *Sci Rep*. (2017) 7:12261. doi: 10.1038/s41598-017-12434-z

57. Chou WY, Cheng YS, Ho CL, Liu ST, Liu PY, Kuo CC, et al. Human spot 14 protein interacts physically and functionally with the thyroid receptor. *Biochem Biophys Res Commun*. (2007) 357:133–8. doi: 10.1016/j.bbrc.2007.03.103

58. Chou WY, Ho CL, Tseng ML, Liu ST, Yen LC, Huang SM. Human spot 14 protein is a P53-dependent transcriptional coactivator via the recruitment of thyroid receptor and ZAC1. *Int J Biochem Cell Biol*. (2008) 40:1826–34. doi: 10.1016/j.biocel.2008.01.014

59. Dumas SN, Guo CA, Kim JK, Friedline RH, Ntambi JM. Interleukin-6 derived from cutaneous deficiency of Stearoyl-CoA desaturase-1 may mediate metabolic organ crosstalk among skin, adipose tissue and liver. *Biochem Biophys Res Commun*. (2019) 508:87–91. doi: 10.1016/j.bbrc.2018.11.083

60. Zappaterra M, Deserti M, Mazza R, Braglia S, Zambonelli P, Davoli R. A gene and Protein expression study on four porcine genes related to intramuscular fat deposition. *Meat Sci*. (2016) 121:27–32. doi: 10.1016/j.meatsci.2016.05.007

61. Smith SB, Kawachi H, Choi CB, Choi CW, Wu G, Sawyer JE. Cellular regulation of bovine intramuscular adipose tissue development and composition. *J Anim Sci*. (2009) 87:E72–82. doi: 10.2527/jas.2008-1340

62. Doran O, Moule SK, Teye GA, Whittington FM, Hallett KG, Wood JD. A reduced protein diet induces Stearoyl-CoA desaturase protein expression in pig muscle but not in subcutaneous adipose tissue: relationship with intramuscular lipid formation. *Br J Nutr*. (2006) 95:609–17. doi: 10.1079/BJN20051526

63. Freake HC, Moon YK. Hormonal and nutritional regulation of Lipogenic enzyme mRNA levels in rat primary white and Brown adipocytes. *J Nutr Sci Vitaminol (Tokyo)*. (2003) 49:40–6. doi: 10.3177/jnsv.49.40

64. Freake HC, Oppenheimer JH. Stimulation of S14 mRNA and lipogenesis in Brown fat by hypothyroidism, cold exposure, and cafeteria feeding: evidence supporting a general role for S14 in lipogenesis and lipogenesis in the maintenance of thermogenesis. *Proc Natl Acad Sci USA*. (1987) 84:3070–4. doi: 10.1073/pnas.84.9.3070

65. Jump DB. Rapid induction of rat liver S14 gene transcription by thyroid hormone. *J Biol Chem*. (1989) 264:4698–703. doi: 10.1016/S0021-9258(18)83799-1

66. Jump DB, Narayan P, Towle H, Oppenheimer JH. Rapid effects of triiodothyronine on hepatic gene expression. Hybridization analysis of tissue-specific triiodothyronine regulation of Mrnas14. *J Biol Chem*. (1984) 259:2789–97. doi: 10.1016/S0021-9258(17)43215-7

67. Wang X, Cheng J, Qin W, Chen H, Chen G, Shang X, et al. Polymorphisms in 5' proximal regulating region of THRSP gene are associated with fat production in pigs. *3 Biotech*. (2020) 10:267. doi: 10.1007/s13205-020-02266-6

68. Wang S, Pan C, Ma X, Yang C, Tang L, Huang J, et al. Identification and functional verification reveals that MiR-195 inhibiting THRSP to affect fat deposition in Xinyang Buffalo. *Front Genet*. (2021) 12:736441. doi: 10.3389/fgene.2021.736441

69. Ahonen MA, Horing M, Nguyen VD, Qadri S, Taskinen JH, Nagaraj M, et al. Insulin-inducible THRSP maintains mitochondrial function and regulates sphingolipid metabolism in human adipocytes. *Mol Med*. (2022) 28:68. doi: 10.1186/s10020-022-00496-3



OPEN ACCESS

EDITED BY

Yalan Yang,
The University of Chicago, United States

REVIEWED BY

Zeying Wang,
Shenyang Agricultural University, China
Shanhe Wang,
Yangzhou University, China

*CORRESPONDENCE

Zhihong Liu
✉ Liuzh7799@163.com

RECEIVED 22 March 2024

ACCEPTED 14 June 2024

PUBLISHED 05 July 2024

CITATION

Zhang C, Qin Q, Liu Z, Wang Y, Lan M, Zhao D, Zhang J, Wang Z, Li J and Liu Z (2024) Combining multiomics to analyze the molecular mechanism of hair follicle cycle change in cashmere goats from Inner Mongolia.
Front. Vet. Sci. 11:1405355.
doi: 10.3389/fvets.2024.1405355

COPYRIGHT

© 2024 Zhang, Qin, Liu, Wang, Lan, Zhao, Zhang, Wang, Li and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Combining multiomics to analyze the molecular mechanism of hair follicle cycle change in cashmere goats from Inner Mongolia

Chongyan Zhang^{1,2}, Qing Qin^{1,3}, Zhichen Liu^{2,3},
Yichuan Wang^{1,2}, Mingxi Lan^{1,2}, Dan Zhao^{2,3}, Jingwen Zhang^{1,2},
Zhixin Wang^{1,3}, Jinquan Li^{1,2,3} and Zhihong Liu^{1,2,3,4*}

¹Department of Animal Science, Inner Mongolia Agricultural University, Hohhot, China, ²Inner Mongolia Key Laboratory of Sheep & Goat Genetics Breeding and Reproduction, Hohhot, China, ³Key Laboratory of Mutton Sheep & Goat Genetics and Breeding, Ministry of Agriculture and Rural Affairs, Hohhot, China, ⁴Northern Agriculture and Livestock Husbandry Technical Innovation Center, Chinese Academy of Agricultural Sciences, Hohhot, China

Sheep body size can directly reflect the growth rates and fattening rates of sheep and is also an important index for measuring the growth performance of meat sheep. Inner Mongolia Cashmere Goat is a local excellent breed of cashmere and meat dual-purpose, which is a typical heterogeneous indumentum. The hair follicles cycle through periods of vigorous growth (anagen), a regression caused by apoptosis (catagen), and relative rest (telogen). At present, it is not clear which genes affect the cycle transformation of hair follicles and unclear how proteins impact the creation and expansion of hair follicles. We using multi-omics joint analysis methodologies to investigated the possible pathways of transformation and apoptosis in goat hair follicles. The results showed that 917, 1,187, and 716 proteins were specifically expressed in anagen, catagen and telogen. The result of gene ontology (GO) annotation showed that differentially expressed proteins (DEPs) are in different growth cycle periods, and enriched GO items are mostly related to the transformation of cells and proteins. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment result indicated that the apoptosis process has a great impact on hair follicle's growth cycle. The results of the protein interaction network of differential proteins showed that the ribosomal protein family (RPL4, RPL8, RPS16, RPS18, RPS2, RPS27A, RPS3) was the core protein in the network. The results of combined transcriptome and proteomics analysis showed that there were 16, 34, and 26 overlapped DEGs and DEPs in the comparison of anagen VS catagen, catagen VS telogen and anagen VS telogen, of which API5 plays an important role in regulating protein and gene expression levels. We focused on API5 and Ribosomal protein and found that API5 affected the apoptosis process of hair follicles, and ribosomal protein was highly expressed in the resting stage of hair follicles. They are both useful as molecular marker candidate genes to study hair follicle growth and apoptosis, and they both have an essential function in the cycle transition process of hair follicles. The results of this study may provide a theoretical basis for further research on the growth and development of hair follicles in Inner Mongolian Cashmere goats.

KEYWORDS

apoptosis, ribosomal protein, hair follicle, telogen, anagen, catagen

1 Introduction

The integument of cashmere goats is separated into two sections: cashmere and wool. The cashmere production process is periodic and is split into three stages: telogen, anagen, and catagen (1, 2). It has been confirmed that the anagen period is from March to September in a year, which is the process of gradual shedding of cashmere in the previous cycle, new hair follicle regeneration, and cashmere growth; From October to December, the hair follicle development stopped in the catagen, and the cashmere continued to grow, in the later stage, the hair follicle structure began to shrink, the cell apoptosis, and the cashmere growth stopped; From January to March, the cashmere enter the telogen, and the hair follicle structure and cashmere remain static (3). This is a complex process regulated by multiple factors. The biggest feature is that there will be obvious gene expression, cell proliferation, and differentiation. Harris (4) studied that IGF-1 is involved in the periodic regulation of hair follicles, specifically by inducing the proliferation and differentiation of cells such as dermal papilla cells and epithelial cells, thereby regulating the periodic circulation of hair follicles. Yano et al. (5) showed that VEGF also plays a role in promoting hair growth in the periodic cycle of hair follicles. Rendl et al. (6) showed that the expression of BMP2 and BMP4 genes in the secondary hair follicles of cashmere goats in the telogen was higher than that in the growth period, which had the effect of inhibiting the transition of hair follicles from the telogen to the growth period and was one of the important factors regulating the periodic growth of cashmere. Although the cycle process of cashmere growth has been extensively verified. However, the molecular regulation mechanism of cashmere entering three stages has not been revealed.

High-throughput sequencing techniques like RNA-Seq can be used to find novel and low-abundance transcripts, allowing researchers to explore the whole range of gene expression and spot transcript variations between samples (7–9). Proteomics is a potent method for revealing the makeup, distribution, alterations, and interactions of proteins in cells, tissues, or organisms. It includes protein and functional patterns (10). Due to the changes in protein abundance, High-abundance proteins with comparable quality or chemical properties can readily obscure low-abundance proteins (11). Mass spectrometry-based comparative proteomics methods, methods like label-free, iTRAQ, and SWATH allow us to pinpoint proteins that exhibit notable alterations in expression levels on a broad scale under particular circumstances (12–14). Proteomics and other omics together provide more valuable information for investigating the cashmere growth and hair follicle development cycle (15). Proteome and transcriptome are two closely connected downstream and upstream genomics (16, 17). Integrating transcriptome and proteome analysis can provide additional insights not available through conventional individual histology and provide a comprehensive understanding of gene expression and regulation at every stage (18).

There, we examined the cashmere growth fluctuations in cashmere goats using the transcriptome and proteome for the first time. We additionally investigated the impact of various periods on cashmere development, transcriptional alterations, and protein levels. Our results fill a gap in the study of cashmere growth and hair follicle development.

2 Materials and methods

2.1 Animal welfare disclaimer

The Yiwei White cashmere Goat Breeding Farm in Erdos, Inner Mongolia, provided grazing conditions for the Inner Mongolian cashmere goats used in this experiment. The Inner Mongolia Agricultural University's experimental animal management committee has authorized every experimental technique used in this work. The present investigation involved the collection of skin samples by the International Guiding Principles for Biomedical Research involving animals. The experiment was approved by the Inner Mongolia Agricultural University's Special Committee on Scientific Research and Academic Ethics, which is in charge of approving the university's biomedical research ethics [Approval No: (2020)056, project title: the International Guiding Principles for Biomedical Research involving animals, approval date: May 6th, 2020]. Three mature cashmere goats at a time, with similar growth, age, and feeding circumstances, serve as samples for three periods. The RNA of the samples is then mixed. Nine cashmere goats from the same family had skin samples taken. In the Department of Surgery, samples were taken from the middle of the scapula at a length of 10–15 cm. The experimental animals were not sacrificed, and after the skin was sampled, they received medication treatments that did not interfere with their normal growth. The 3-cm-diameter skin samples were collected, immediately rinsed with PBS, and quickly frozen in liquid nitrogen. After being transported to the lab in a liquid nitrogen tank, the samples were kept in a freezer at -80°C .

2.2 Protein extraction and digestion

The lysis process involved pulverizing nine skin samples in liquid nitrogen, adding the powder to lysis buffer (8-mg urea and 1% protease inhibitor), then lysing the mixture using ultrasound. The supernatant was then transported to a fresh centrifuge tube after the cell debris was eliminated using a centrifuge set at 12,000 g for 10 min at 4°C . Following the collection of the filtrate, the processed samples were held at -80°C while protein quantification was carried out using the BCA Protein Assay Kit (Abcan, China).

2.3 Mass spectrometry for label-free LC/MS

Peptide data were obtained using information-dependent capture (IDA) and sequential window acquisition of all theoretical spectra-mass spectrometry (SWATH-MS) in the Sciex LC-MS/MS system (Framingham, MA, United States). A C18 column measuring $75\mu\text{m}$ by 15 cm was used to inject about $2\mu\text{g}$ of polypeptide for separation. A linear gradient of 0.1% formic acid in acetonitrile and 0.1% formic acid in water was used to separate the peptides (120 min, 500 nL/min, from 5 to 80%). The IDA parameters were as follows: automated collision energy; 350–1,800 m/z for time-of-flight mass spectrometry collection; 400–1,800 m/z for MS/MS and IDA scan. The nominal resolution was set at 30,000. The following were the SWATH-MS conditions: nominal resolutions of MS1 and MS2, 30,000 and 15,000, respectively; 150–1,200 m/z, MS1 mass range; 100–1,500 m/z, MS2 mass range.

2.4 Label-free LC/MS quantitative and qualitative and profiling

Peptide identification was performed using the UniProt/SWISS-PROT/*Capra hircus* database¹ and Protein Pilot v4.5 software (Sciex, Framingham, MA, United States). A 1% false discovery rate (FDR) was used to filter the results. Trypsin was chosen as the enzyme and two missed cleavage sites were permitted among the search parameters. There was a 15 ppm peptide mass tolerance and a 20 mmu fragment mass tolerance. The data were imported into the software PeakView v2.1 (Sciex, Framingham, MA, United States), and the SWATH database was searched using the ion library produced by Protein Pilot. PeakView produced the extracted ion chromatograms (XICs) by processing the target and nontarget data. The findings were then explained and subjected to a quantitative analysis using the MarkerView v3.0 program (Sciex). With MarkerView, one may quickly examine data to identify the proteins that are differentially expressed (DEPs). The fold change analysis and t-tests were merged into principal component analysis (PCA) and volcano plot analysis. DEPs were identified using a fold change >2 or fold change <0.5, as well as statistical significance ($p < 0.05$).

2.5 Total RNA extraction from skin and construction of sequencing library

Using an RNAiso Plus Kit (TRIzol technique), total RNA was extracted from the skin of three cashmere (one period) goats. The RNA samples from three goats were combined after the total RNA was examined for purity and integrity using a sterile UV-vis spectrophotometer and an Agilent 2,100 bioanalyzer, respectively. RNA samples were collected in September (anagen), December (catagen), and March (telogen) using the same method and operation. The whole RNA was kept at -80°C in a freezer.

The Illumina TruSeq™ RNA Sample Preparation Kit's operating instructions were followed while creating the cDNA library for transcriptome sequencing. By combining equal parts of each, the total RNA from the three cashmere goats was combined. After the mRNA was separated into 100–400 bp mRNA, it was purified using oligo-dT magnetic beads. Exonucleases, polymerases, and fragmented mRNA were used to create double-stranded cDNA. Using a Bio-Rad Certified Low-Range Ultra Agarose Kit, the ends of the double-stranded cDNA fragments were blunted, and the double-stranded cDNAs were phosphorylated to ligate the sequencing adapters and poly (A) tail. The sizes of the cDNA recovered were determined to be 200–300 bp. A sequencing library was created by PCR amplification of cDNA, and a TBS-380 device was used for library quality verification. An Illumina HiSeq™ 2000 sequencing platform was used to perform paired-end sequencing of the cDNA. The samples were sequenced by Beijing Baimaike Biotechnology Co., Ltd. using an A2 × 100 bp sequencing test.

2.6 Bioinformatics analysis

Using the *Capra hircus* genome annotation as background and the David database² with default parameters, GO functional and KEGG pathway annotation was carried out. Obtained, encompassing analyses of the cellular component (CC), molecular function (MF), and biological process (BP).

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)³ database was used to perform the Protein-protein Interaction (PPI) study of the differentially expressed proteins between goat breeds, as previously mentioned. A minimum needed interaction score of >0.4 was used in the construction of the network. The network was visualized using Cytoscape v3.9.0 software (Cytoscape Consortium, San Diego, CA, United States). Additionally, the hub proteins of the PPI network were investigated using the McCreight (MCC) approach with the help of the Cytoscape add-on CytoHubba.

2.7 Parallel reaction monitoring validation for differentially expressed proteins

The protein abundances were examined using the parallel reaction monitoring (PRM) approach to verify the accuracy of the SWATH-based proteomic data. The peptides were separated using a liquid chromatography-tandem quadrupole mass spectrometry system and dissolved in liquid chromatography mobile phase A (0.1% formic acid solution). The final step was processing the generated MS data with Skyline (64-bit, 22.2.0.351) (19).

2.8 Quantitative real-time PCR

For quantitative reverse transcription PCR (qPCR) analysis, the cDNA that was previously acquired was utilized. Primer 3.0 software was utilized to create the gene-specific primers for q-PCR, and Sangon Biotech Co., Ltd. (Shanghai, China) synthesized them. Table 1 lists the fragment sizes and primer sequences. For the q-PCR, a 20-μL reaction volume was used together with a PrimeScript RT Reagent Kit (TaKaRa, Beijing), 10 μL of 2 × SYBR Premix Ex Taq II (TaKaRa), 2 μL of cDNA, and 0.5 μL of each primer. A Bio-Rad IQ5 multicolor real-time PCR detection system (Hercules, CA, United States) was used to evaluate the reaction. The reference used was the β-actin. The $2^{-\Delta\Delta\text{CT}}$ method was used for the qRT-PCR study, and SPSS software (version 17.0) was used for the statistical analysis. The format for values is mean ± standard deviation.

3 Results

3.1 Identification of proteins in goat skin

Cashmere goat skin proteome analysis was carried out in three stages using the constructed proteome map liquid

¹ <https://www.UniProt.org/#>

² <https://david.ncicrf.gov/>

³ <https://cn.string-db.org/>

TABLE 1 Primer sequence information of differentially expressed genes.

Primer name	Primer sequence (5'–3')	Product length/bp
VEGFA	F: TACCACCACCACCACCACCATC R: CACCCGCCCATGAATGCTTCTG	133
RAC1	F: CACTGTCCCAACACCCATCATC R: GGCCTCAGCTTCTTCTCCTTCAG	98
NOTCH1	F: GCAAGTGCATCAACACGCTG R: TGGAAGTCCCGATCTGGTC	145
LHX2	F: CTACTACAATGGCGTGGGCA R: ACTTCATGGTCCGAAGCTGG	212
KRT2	F: TCTGCAGCCTCTCAATGTGA R: GGTGCCACATCGATTTGTT	184
FGF21	F: TGCTGGCTGTCTCTCTGCTAG R: TCCTGGGCATCATCCGTGTAGAG	118bp

chromatography–tandem mass spectrometry approach to examine changes in protein dynamics during the growth of goat wool. The proteome and transcriptome profiling of inner mongolia cashmere goat skin with anagen (Sept), catagen (Dec), and telogen (March) were constructed. Totally, with a false discovery rate (FDR) ≤ 0.01 , 1890 proteins and 20,116 distinct peptides were identified in March; 2,460 proteins and 26,191 distinct peptides were identified in September; 2,505 proteins and 27,352 distinct peptides were identified in December. 716, 917, and 1,187 proteins were specifically expressed in March, September, and December (Figure 1A). Among three groups, a total of 2,983 proteins and 38,569 distinct peptides were identified (Figure 1B). We used principal component analysis (PCA) to cluster the protein to evaluate the quantitative consistency of the proteomics data. (Figure 1C). The PCA result showed that most of the samples in the same period were clustered together and crossed in the quadrants of different periods, indicating that proteins were related to different periods of hair follicle development. According to Figure 1G, it can be found that the number of peptides and maps did not change significantly during the transition from the hair follicle growth period to the degenerative period, that is, the basic composition of these proteins did not change. However, in this process, the number of proteins has changed significantly, and the change rule is that the number of proteins in the growth period-degeneration period-rest period changes from more to less and then to more. The number of proteins gradually increased during the growth period to the degenerative period, revealing the important role of these proteins in the development of hair follicles and the growth of cashmere. However, there are two primary phases to the growth of hair follicles, the first is the growth of primary hair follicles, and the second is the creation of cashmere fiber. Different proteins are involved in these two processes, but the total protein number is gradually increasing. Previous studies have shown evidence for the presence of molecular spatial gradients during the formation of hair follicles. The hair follicle structure as a function of the spatial dimension was then examined in our IDA dataset to capture this at the proteome level. The study used IDA data with fold changes ≥ 2 or < 0.05 and a p -value < 0.05 to identify 137 proteins that were found to be highly expressed in anagen compared with telogen, while 656 proteins were highly expressed in telogen compared with anagen (Figure 1D). In anagen VS catagen, 644 up-regulated proteins and 30 down-regulated proteins were identified

(Figure 1E). In catagen VS telogen 355 up-regulated proteins and 362 down-regulated proteins were identified (Figure 1F). These proteins play different roles in three stages to promote hair follicle growth and apoptosis.

3.2 Functional enrichment analysis of differentially expressed proteins

To explore the potential functions of up-regulated and down-regulated Differentially Expressed Proteins (DEPs) identified in this study, GO and KEGG enrichment analysis was performed on DEPs. The results of GO enrichment in three different periods showed that all categories were enriched, and the biological process items were mainly enriched. The main GO terms of biological processes in the anagen VS catagen groups include translation, positive regulation of cell proliferation, protein stabilization, intracellular protein transport, protein transport, vesicle-mediated transport, etc. (Figure 2A). It shows that 674 differential proteins play an important role in promoting cashmere growth by hair follicles. In the catagen VS telogen comparison group, protein transport, skin morphogenesis, aging, cell death, positive regulation of macroautophagy, etc. were enriched (Figure 2B), indicating that 717 differential proteins were mainly used in hair follicle apoptosis. In the telogen VS anagen comparison group, translation, intracellular protein transport, cell differentiation, ubiquitin-dependent ERAD pathway, regulation of translational initiation, positive regulation of protein import into the nucleus etc. were enriched (Figure 2C), it is indicated that the role of the up-regulated protein in the transition from telogen to anagen is hair follicle apoptosis, while the role of the down-regulated protein is the growth or maintenance of cashmere.

We performed KEGG enrichment analysis on all DEPs. Within these three divisions, DEPs were mainly enriched in Metabolic pathways, Oxidative phosphorylation, and Estrogen signaling pathways. For the anagen VS catagen group, DEPs were significantly enriched in Phagosome, Proteasome, Protein digestion and absorption, Apoptosis, mTOR signaling pathway, and PPAR signaling pathway (Figure 2A). For the catagen VS telogen group, DEPs were significantly enriched in Pathways of neurodegeneration-multiple diseases, AMPK signaling pathway, Glutathione metabolism, and Leukocyte transendothelial migration (Figure 2B). For the telogen VS anagen group, Protein processing in the endoplasmic reticulum, Glutathione metabolism, Biosynthesis of amino acids were significantly enriched (Figure 2C). From the results of different pathways enriched from different periods, it is further verified that different proteins are enriched and functioned at different periods to maintain the cycle of cashmere and the cycle of cashmere growth.

Protein–protein interaction (PPI) network utilizing STRING was created to ascertain the relationships among these DEPs. The PPI network of DEPs contains 11,621 edges and 1,064 nodes. Additionally, the top 50 proteins in the network were obtained based on the analysis using Cytoscape Degree (Figure 2D). Ribosomal protein family (RPL4, RPL8, RPS16, RPS18, RPS2, RPS27A, RPS3), Elongation Factor family (EEF1G, EEF2, EFTUD2), Translation Initiation Factor family (EIF2S1) and Glutamyl-prolyl-tRNA synthetase 1 (EPRS1) interacted strongly with each other. In addition, we found these proteins were significantly enriched in translation, cell differentiation, positive regulation of cell proliferation, negative regulation of translation, etc. based on PPI and function enrichment results. We speculated that the ribosomal protein

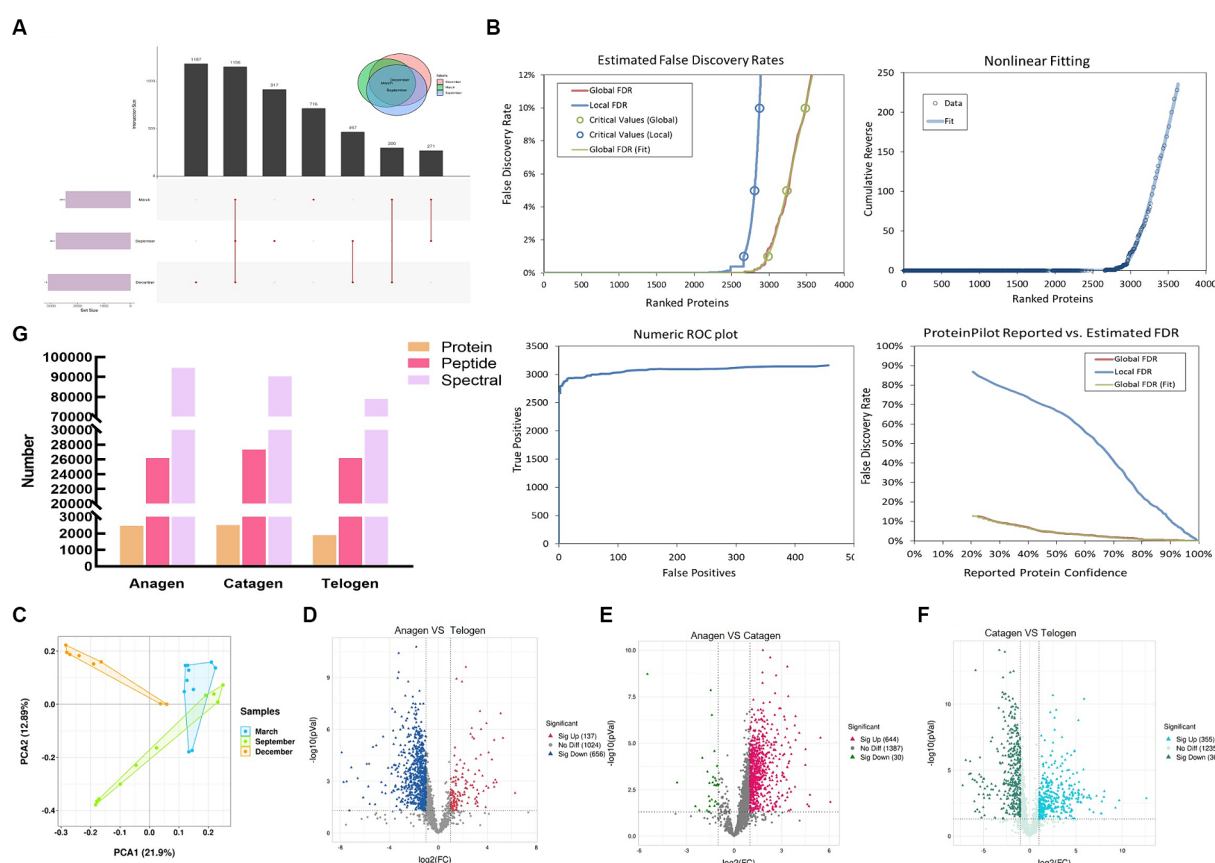


FIGURE 1

Identification of proteins of skins from goats. (A) UpSet plots depicting the number of unique and shared expressed proteins between different periods. (B) Estimated False Discovery Rates, Nonlinear Fitting of, Numeric ROC plot, and ProteinPilot Reported vs. Estimated FDR of all protein. (C) Two-dimensional scatter plot of quantitative principal component analysis of protein among samples. (D) Volcano map of differentially expressed proteins in anagen and telogen. (E) Volcano map of differentially expressed proteins in anagen and catagen. (F) Volcano map of differentially expressed proteins in catagen and telogen.

might be involved in development regulation, cell differentiation, and apoptotic in hair follicles of cashmere goats at different stages.

3.3 Analysis of differences in gene expression

To explore the molecular mechanism of initiating periodic changes in hair follicles at the genomic level. Three skin samples from cashmere goats aged 3, 9, and 12 months were subjected to transcriptome sequencing analysis using the Illumina sequencing technology. The raw read count that was retrieved was 173,925,724. Over 94% of bases with a mass value greater than 30 (and an error rate of less than 0.1%) were found in all sequenced samples. 169,651,142 clean readings were obtained thereafter data filtering. The retrieved 220,696,075 mapped reads had a comparability rate of 92.58–93.22% with the reference genome of cashmere goats (Table 2).

By comparing the levels of gene expression in each sample, Differentially Expressed Genes (DEGs) were identified. Totally, 775, 882, and 1,034 DEGs were determined in comparison of anagen VS catagen, catagen VS telogen, and telogen VS anagen ($q \leq 0.05$, $|\log_2(FC)| > 2$). Furthermore, 58 DEGs that overlapped were found in each of the three comparison groups (Figure 3A). Cluster analysis of DEGs

found that most of the DEGs in the anagen were down-regulated, most of the DEGs in the telogen were up-regulated, and the DEGs in the catagen period were in between (Figure 3B).

3.4 Integrated analysis of transcriptome and proteome

The DEPs and DEGs were compared within each group to assess the correlation between the transcriptome and proteome data. Three clusters of genes with 16, 34, and 26 overlapped copies were found (Figures 4A–C).

Translational regulatory mechanisms explain the correlation between transcriptome and proteome data. At this point, nine-quadrant association analyses can be very helpful. Among the 16 genes/proteins expressed in anagen VS catagen, TCHH was enriched in the 1 quadrant; in the 2 quadrants, COL2A1 is enriched; in the 4 quadrant, API5, CDSN, FBN1, TPPP3, UCHL3 were enriched. HSPA8, LDHB, OGN, PON3, RPS24, SBSN, and SOAT1 were enriched in the 5 quadrants. In the 6 quadrants, GNG12 and CCT2 were enriched (Figure 4D). Of the 34 genes/proteins expressed in catagen VS telogen, COCH was enriched in the 1 quadrant; APOA1, COL1A2, EIF4A2, MYH11, and RPL27A were enriched in the 2 quadrant; ALDOC, API5, KRT10, SLC25A24, TGFBI,

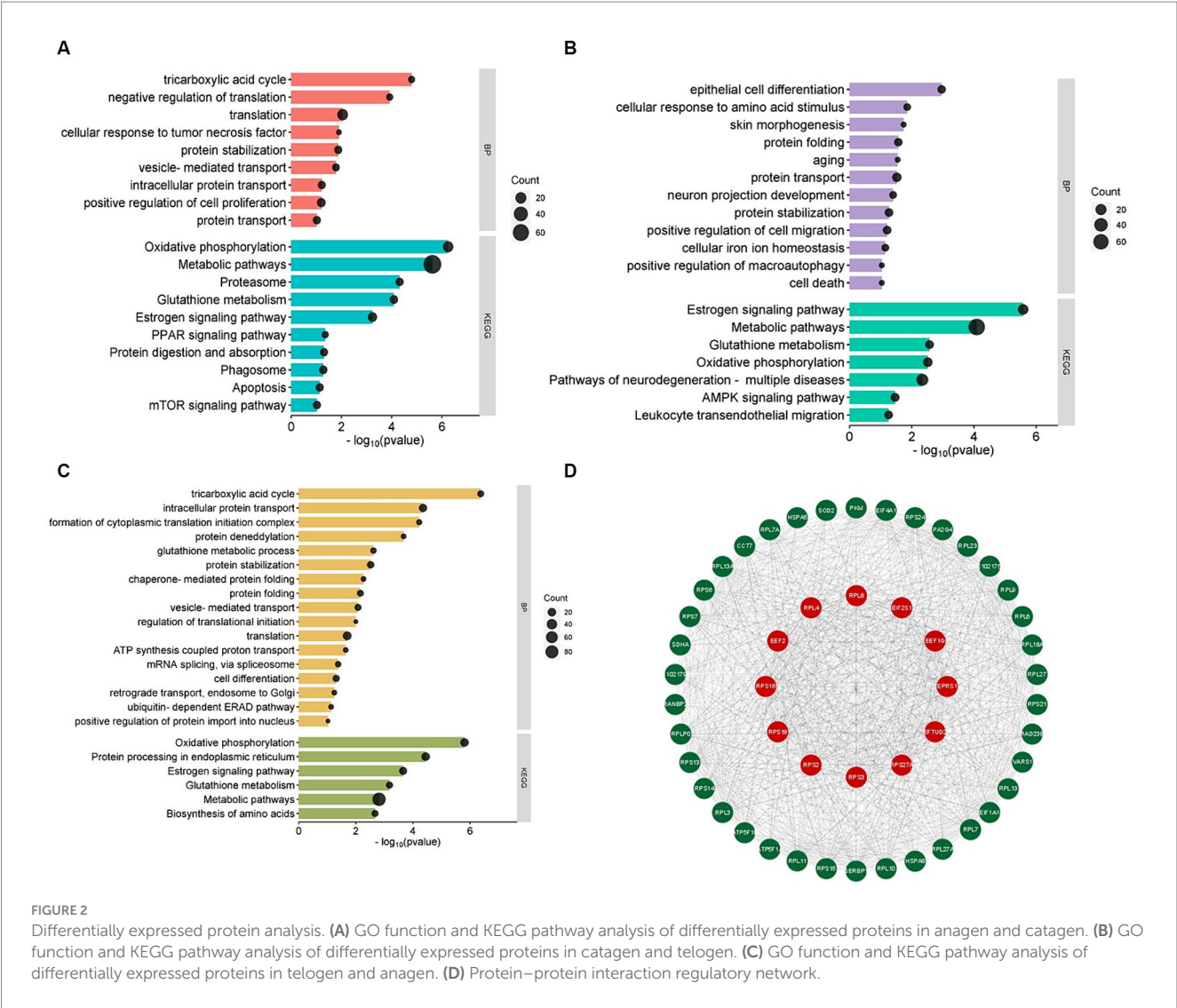


TABLE 2 The mapping of clean data.

Library	Raw reads	Clean reads	GC (%)	N (%)	Q30 (%)
March	83,697,110	81,577,244	53.07	0.0111	94.54
September	82,490,126	80,504,704	53.28	0.0113	94.57
December	7,738,488	7,569,194	52.98	0.0112	94.5

and *UCHL3* were enriched in the 4 quadrants; *FBLN5*, *LDHBC4*, *AHCY*, *DCN*, *DNAJA2*, *DPT*, *FBN1*, *GANAB*, *LUM*, *MMP2*, *PADI3*, *QSOX1*, *SLC1A5*, and *TES* were enriched in the 5 quadrants; *AdipoQ*, *GNG12*, *PLIN3*, *PRDX4*, and *PSMC4* were enriched in the 6 quadrants; *PCP4L1* and *CDSN* were enriched in the 7 and 8 quadrants (Figure 4E); The 26 genes/ proteins expressed in telogen VS anagen, *RPL23* was enriched in the 2 quadrantss; *DNAJA2*, *DPT*, *GPX3*, *PCP4L1*, *RPL21*, *RPL22*, and *SLC25A24* were enriched in the 4 quadrant; *ACTG2*, *COCH*, *EIF4A2*, *GLUL*, *GSR*, *GSS*, *HSPH1*, *MAOA*, *PMM2*, *RPL10*, *RPL4*, *TPM2*, *TPPP3*, *TXN2*, and *UCHL3* were enriched in the 5 quadrant; *API5* an lation-level regulation may be the reason why certain proteins abundant in the sixth, eighth, and ninth quadrants had higher abundances than RNA, whereas some proteins in the first, second, and fourth quadrants had

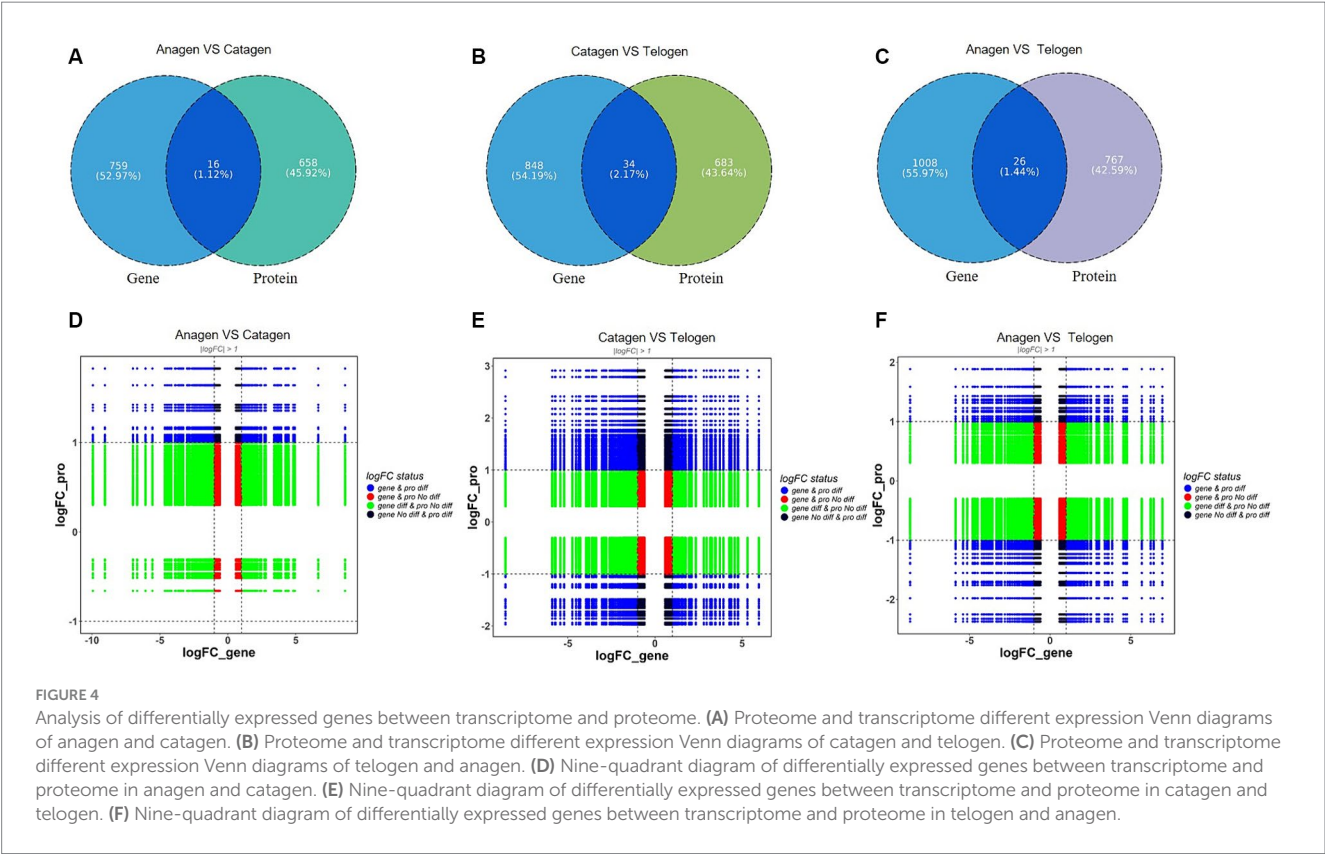
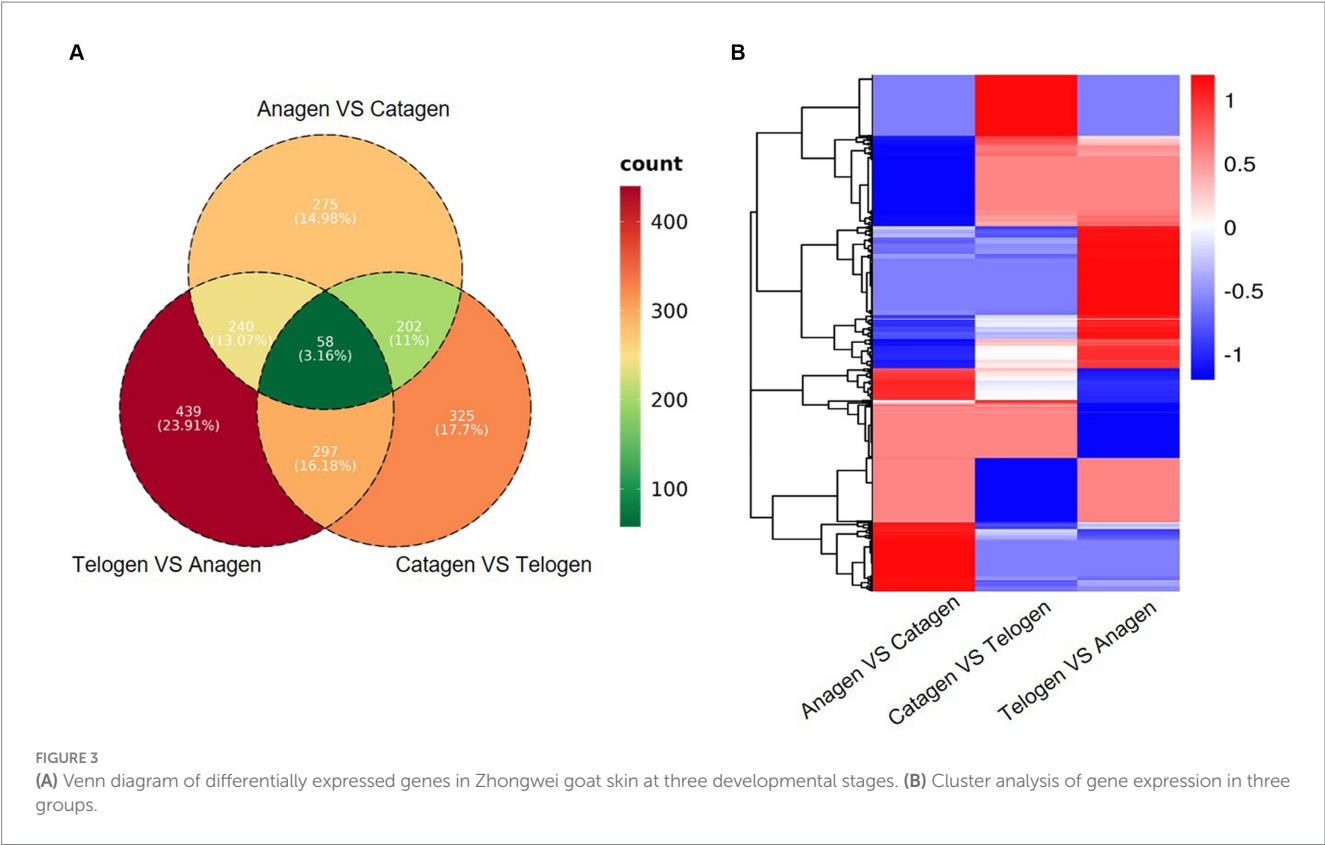
lower abundances than the comparable RNAs. From the anagen-catagen and the catagen-telogen, both are expressed by *API5*; the high expression of *RPL* family members was found in the process of telogen-anagen.

3.5 Key protein structure domain analysis

The main function of the gene is to store and transmit genetic information, through the process of transcription and translation, and ultimately the formation of protein, and then play a role in the body. Proteins account for 50% of the dry mass of cells and play a role in everything the organism does, they are all composed of the same 20 amino acids. The *API5* sequence length of 504, 56.1% is Helix, 1.2% is Strand, and 42.7% is non-helical and non-folded random structure. On Solvent Accessibility, 42.1% of the area is Buried and 48.7% is Exposed (Figure 5).

3.6 PRM quantitative results

To confirm the correctness of the proteomic data in three comparison groups, three proteins were chosen for PRM quantification



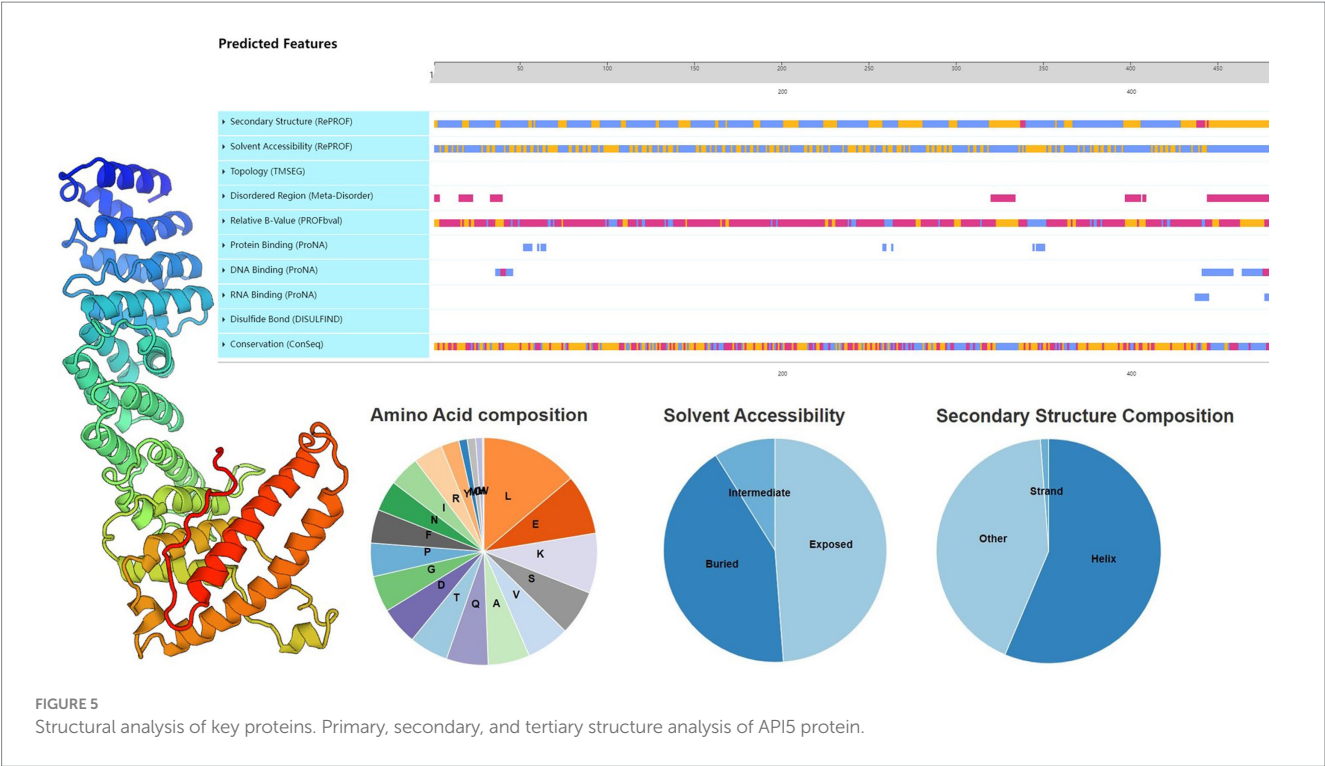


TABLE 3 PRM result compared with SWATH-based quantitative result.

Gene name	Peptide	Telogen expression (PRM)	Telogen expression (SWATH)	Anagen expression (PRM)	Anagen expression (SWATH)	Catagen expression (PRM)	Catagen expression (SWATH)
TPM4	RFEKPLEEK	4.57E+04	18.12	9.93E+04	22	3.31E+05	31.86
SNX2	VSHYIINSLPNRRFK	2.34E+03	4.11	2.48E+05	10.84	7.41E+03	4.29
PSMA1	IGGAQNRSYSK	2.15E+05	8.2	2.85E+05	11.7	2.44E+04	6.17

based on the expression level of the DEPs. The findings of the SWATH data analysis and PRM detection ratio showed a steady general trend, suggesting that the proteomic data were dependable and repeatable (Table 3).

3.7 Quantitative real-time PCR validation

Six DEGs were chosen at random to verify the correctness of the transcriptome data (Figure 6). Since the transcriptome results matched the DEGs’ real-time fluorescence expression patterns, our transcriptomic data were considered reliable.

4 Discussion

From phenotypic to molecular processes, researchers have attempted to investigate the periodic changes in cashmere goat hair follicles since the 1960s. Dynamic changes in gene regulation and protein expression occur during this process (20). Since gene expression at the RNA and protein levels varies with time and tissue, transcriptome sequencing provides a direct means of investigating changes in gene expression, while proteome identification provides a

direct means of investigating changes in protein function (21). We used multi-omics technology to jointly analyze the cycle conversion mechanism of cashmere goat hair follicles.

Firstly, the protein expression at different stages was identified. As the hair follicle began to grow and develop to the hatching hair grew out of the body surface and then to the hair follicle atrophy and apoptosis, the number of proteins was consistent with its changes. We speculate that there are specific proteins that function at different stages and promote the growth and apoptosis of hair follicles. The differential proteins were analyzed, and the enriched functional results including cell proliferation, protein transport cell death, etc., were as we speculated. Secondly, we performed gene mining on skin samples at different stages through transcriptomics. It was found that, like proteomics, the number of gene expressions fluctuated with the growth and apoptosis of hair follicles, and specific expression genes appeared at different stages. Among the 2,691 DEGs identified, 76 DEGs were overlapped with DEPs. Subsequently, we focused on the genes/proteins of 76 and finally locked the Nuclear protein Apoptosis Inhibitor 5 and ribosomal protein (RP) through correlation analysis.

Nuclear protein Apoptosis Inhibitor 5 (API5) prevents cells from going through the apoptotic process. This protein was first discovered in serum-deprived cells that survived, and it was subsequently discovered to be overexpressed in some malignancies and to control

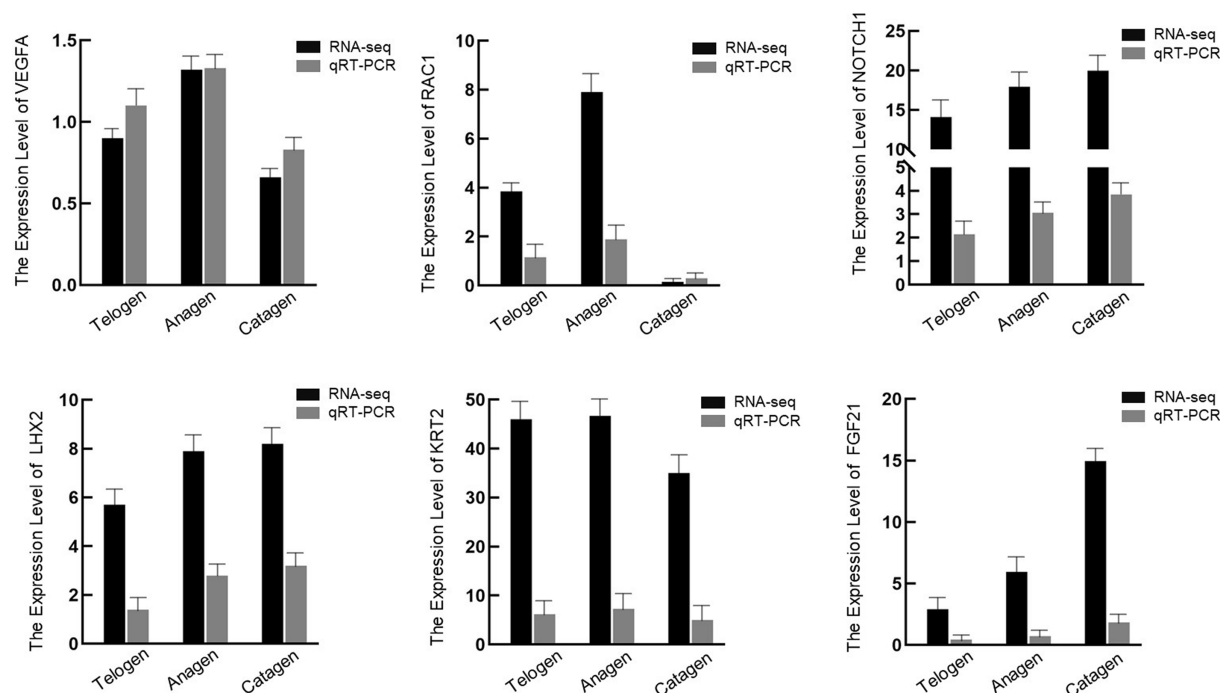


FIGURE 6

mRNA expression levels for six genes in the skin samples of different period cashmere goats examined, via quantitative polymerase chain reaction, to verify the RNA sequencing data.

apoptosis in both vertebrates and invertebrates (22–24). In our study, we found that the API5 protein is regulated by post-transcriptional and translational levels, and the target gene inhibits its translation during the anagen-catagen and catagen-telogen periods. In these two stages, hair follicle cells undergo a process from initial apoptosis to complete apoptosis, and the expression of API5 shows a continuous downward trend, which is consistent with the cycle transition of goat hair follicles. During the telogen-anagen period, the API5 protein is regulated by post-transcriptional and translation levels, and the target gene promotes its translation. At this stage, hair follicles begin to grow and develop, API5 functions to inhibit apoptosis, and API5 expression increases; Numerous studies have demonstrated that API5 is up-regulated in human cancers of the cervix, prostate, lung, colon, and other related tissues in mouse fibroblasts, and has significant anti-apoptotic ability (22, 24–26). However, there is no relevant report on the skin cycle of goats. In our study, API5 protein showed a trend of increasing first and decreasing in the process of telogen-growth-degeneration, and the gene expression pattern was the same trend. For the first time, we discovered that in goats, API5 is connected to both apoptosis and hair follicle growth. Combined with functional analysis, API5 was enriched in Apoptosis in the biological process, which further confirmed the important role of API5 in the expression of hair follicle cycle transition in goats. Therefore, API5 will also be the focus of the study of the skin cycle.

Ribosomal protein (RP) is a general term for all proteins involved in the formation of ribosomes. It has a significant role in intracellular protein biosynthesis and has been found in many organisms. It is expressed in different organs and tissues of the human body. Currently, eukaryotic cells contain about 80 different types of ribosomal protein (RP). These RPs are referred to as ribosomal protein large (RPL) and

ribosomal protein small (RPS) based on where they originate from—big and small subunits, respectively. A variety of studies have found that RP not only has the function of protein translation but also participates in the regulation of DNA replication, transcription, repair, RNA splicing, modification, cell proliferation, apoptosis, and other functions (27–29). In this study, RPL family members are the most connected proteins in the protein interaction network. In the co-expression module of DEGs and DEPs, they are highly expressed in the telogen-growth period. This may mean that the ribosomal protein caused a significant shift in the internal molecular microstate of the epidermis when the hair follicle transitioned from the resting phase to the new round of cycle growth phase.

According to our research, ribosomal protein and apoptosis inhibitor 5 had an impact on the growth and development of hair follicles in Inner Mongolian Cashmere goats. This will support our ongoing research into the mechanisms behind the proliferation and death of hair follicles in Inner Mongolian cashmere goats.

5 Conclusion

We provide comprehensive proteomic and transcriptomic data for inner mongolia cashmere goat skin with different three development stages. The gene expression at different stages and the function of protein volatilization were systematically studied. The data revealed the API5 may be involved in hair follicle apoptosis and influence the process of hair follicle cycle transition, and the transition of hair follicles from telogen to anagen is facilitated by the high expression of the RPL family, which also stimulates hair growth. The results provide a useful transcriptomic and proteomic resource and a broad

understanding of the proliferation apoptosis mechanism underlying hair follicles.

Data availability statement

The data presented in the study are deposited in the ProteomeXchange Consortium repository, accession number PXD036685.

Ethics statement

Samples were collected by the Guidelines for Experimental Animals of the Ministry of Science and Technology (Beijing, China) and were approved by the experimental animal ethics committee of Inner Mongolia Agricultural University (GB 14925–2001).

Author contributions

CZ: Writing – original draft, Writing – review & editing. QQ: Writing – review & editing. ZhicL: Writing – review & editing. YW: Writing – review & editing. ML: Writing – review & editing. DZ: Writing – review & editing. JZ: Writing – review & editing. ZW: Writing – review & editing. JL: Writing – review & editing. ZhihL: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work received funding from the National Natural Science Foundation of China(32060742), the National Key R&D Program of China(2021YFD1200901), the Project of Northern Agriculture and

Livestock Husbandry Technical Innovation Center, Chinese Academy of Agricultural Sciences(BFGJ2022002), the Major science and technology projects of Inner Mongolia Autonomous Region (2020ZD0004), University Youth Science and Technology Talent Support Project of Inner Mongolia Autonomous Region(NJYT23012), the central government guides local science and technology development funds(2022ZY0185), Higher educational Basic scientific research business fee project of Inner Mongolia Autonomous Region(BR221005), Supported by Program for Innovative Research Team in Universities of Inner Mongolia Autonomous Region(NMGIRT2322), Xinjiang Key Laboratory of Animal Biotechnology open project funded project “Kazakh sheep body size detection model optimization and weight estimation research”(2023SWS004).

Acknowledgments

The authors are grateful for the samples provided by the Farm. ZhihL provided an experimental platform.

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

- Ge W, Wang SH, Sun B, Zhang YL, Shen W, Khatib H, et al. Melatonin promotes cashmere goat (*Capra hircus*) secondary hair follicle growth: a view from integrated analysis of long non-coding and coding RNAs. *Cell Cycle*. (2018) 17:1255–67. doi: 10.1080/15384101.2018.1471318
- Geng R, Yuan C, Chen Y. Exploring differentially expressed genes by RNA-seq in cashmere goat (*Capra hircus*) skin during hair follicle development and cycling. *PLoS One*. (2013) 8:e62704. doi: 10.1371/journal.pone.0062704
- Liu Z, Yang F, Zhao M, Ma L, Li H, Xie Y, et al. The intragenic mRNA-microRNA regulatory network during Telogen-Anagen hair follicle transition in the cashmere goat. *Sci Rep*. (2018) 8:14227. doi: 10.1038/s41598-018-31986-2
- Harris PM, McBride BW, Gurnsey MP, Sinclair BR, Lee J. Direct infusion of a variant of insulin-like growth factor-I into the skin of sheep and effects on local blood flow, amino acid utilization and cell replication. *J Endocrinol*. (1993) 139:463–72. doi: 10.1677/joe.0.1390463
- Yano K, Brown LF, Detmar M. Control of hair growth and follicle size by Vegf-mediated angiogenesis. *J Clin Invest*. (2001) 107:409–17. doi: 10.1172/jci11317
- Rendl M, Polak L, Fuchs E. Bmp signaling in dermal papilla cells is required for their hair follicle-inductive properties. *Genes Dev*. (2008) 22:543–57. doi: 10.1101/gad.1614408
- Wang L, Cai B, Zhou S, Zhu H, Qu L, Wang X, et al. RNA-seq reveals transcriptome changes in goats following myostatin gene knockout. *PLoS One*. (2017) 12:e0187966. doi: 10.1371/journal.pone.0187966
- Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for Transcriptomics. *Nat Rev Genet*. (2009) 10:57–63. doi: 10.1038/nrg2484
- Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. (2011) 12:87–98. doi: 10.1038/nrg2934
- Pandey A, Mann M. Proteomics to study genes and genomes. *Nature*. (2000) 405:837–46. doi: 10.1038/35015709
- Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*. (1996) 13:19–50. doi: 10.1080/02648725.1996.10647923
- Marcelino I, Almeida AM, Brito C, Meyer DF, Barreto M, Sheikboudou C, et al. Proteomic analyses of *Ehrlichia ruminantium* highlight differential expression of Map1-family proteins. *Vet Microbiol*. (2012) 156:305. doi: 10.1016/j.vetmic.2011.11.022
- Ouyang H, Wang Z, Chen X, Yu J, Li Z, Nie Q. Proteomic analysis of chicken skeletal muscle during embryonic development. *Front Physiol*. (2017) 8:281. doi: 10.3389/fphys.2017.00281
- Murgiano L, D'Alessandro A, Egidi MG, Crisà A, Prosperini G, Timperio AM, et al. Proteomics and Transcriptomics investigation on longissimus muscles in large white and Casertana pig breeds. *J Proteome Res*. (2010) 9:6450–66. doi: 10.1021/pr100693h
- Bourmaud A, Gallien S, Domon B. Parallel reaction monitoring using quadrupole-orbitrap mass spectrometer: principle and applications. *Proteomics*. (2016) 16:2146–59. doi: 10.1002/pmic.201500543
- Hu Y, Zhang Y, Šmarda P, Bureš P, Guo Q. Transcriptome and proteome associated analysis of flavonoid metabolism in haploid *Ginkgo biloba*. *Int J Biol Macromol*. (2023) 224:306–18. doi: 10.1016/j.ijbiomac.2022.10.125

17. Zhao J, Yang J, Xie Y. Improvement strategies for the Oral bioavailability of poorly water-soluble flavonoids: an overview. *Int J Pharm.* (2019) 570:118642. doi: 10.1016/j.ijpharm.2019.118642
18. Li Z, Liu N, Zhang W, Wu C, Jiang Y, Ma J, et al. Integrated transcriptome and proteome analysis provides insight into chilling-induced dormancy breaking in *Chimonanthus praecox*. *Hortic Res.* (2020) 7:198. doi: 10.1038/s41438-020-00421-x
19. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* (2010) 26:966–8. doi: 10.1093/bioinformatics/btq054
20. Gat U, DasGupta R, Degenstein L, Fuchs E. *De novo* hair follicle morphogenesis and hair tumors in mice expressing a truncated Beta-catenin in skin. *Cell.* (1998) 95:605–14. doi: 10.1016/s0092-8674(00)81631-1
21. Su R, Gong G, Zhang L, Yan X, Wang F, Zhang L, et al. Screening the key genes of hair follicle growth cycle in inner Mongolian cashmere goat based on RNA sequencing. *Arch Anim Breed.* (2020) 63:155–64. doi: 10.5194/aab-63-155-2020
22. Krejci P, Pejchalova K, Rosenbloom BE, Rosenfelt FP, Tran EL, Laurell H, et al. The antiapoptotic protein Api5 and its partner, high molecular weight FGF2, are up-regulated in B cell chronic lymphoid leukemia. *J Leukoc Biol.* (2007) 82:1363–4. doi: 10.1189/jlb.0607425
23. Tewari M, Yu M, Ross B, Dean C, Giordano A, Rubin R. AAC-11, a novel cDNA that inhibits apoptosis after growth factor withdrawal. *Cancer Res.* (1997) 57:4063–9.
24. Sasaki H, Moriyama S, Yukiue H, Kobayashi Y, Nakashima Y, Kaji M, et al. Expression of the antiapoptosis gene, AAC-11, as a prognosis marker in non-small cell lung cancer. *Lung Cancer.* (2001) 34:53–7. doi: 10.1016/s0169-5002(01)00213-6
25. Kim JW, Cho HS, Kim JH, Hur SY, Kim TE, Lee JM, et al. Aac-11 overexpression induces invasion and protects cervical cancer cells from apoptosis. *Lab Invest.* (2000) 80:587–94. doi: 10.1038/labinvest.3780063
26. Van den Berghe L, Laurell H, Huez I, Zanibellato C, Prats H, Bugler B. Fgf-2 (fibroblast growth factor-2) (FGF-2)-interacting-factor, a nuclear putatively antiapoptotic factor, interacts specifically with Fgf-2. *Mol Endocrinol.* (2000) 14:1709–24. doi: 10.1210/mend.14.11.0556
27. Wier EM, Neighoff J, Sun X, Fu K, Wan F. Identification of an N-terminal truncation of the NF-Kb P65 subunit that specifically modulates ribosomal protein S3-dependent NF-Kb gene expression. *J Biol Chem.* (2012) 287:43019–29. doi: 10.1074/jbc.M112.388694
28. Seong KM, Jung SO, Kim HD, Kim HJ, Jung YJ, Choi SY, et al. Yeast ribosomal protein S3 possesses a B-Lyase activity on damaged DNA. *FEBS Lett.* (2012) 586:356–61. doi: 10.1016/j.febslet.2011.12.030
29. Pan R, Zhu M, Yu C, Lv J, Guo Y, Bian Z, et al. Cancer incidence and mortality: a cohort study in China, 2008–2013. *Int J Cancer.* (2017) 141:1315–23. doi: 10.1002/ijc.30825



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Liu Yang,
Agricultural Research Service (USDA),
United States
Guoliang Li,
Huazhong Agricultural University, China

*CORRESPONDENCE

Linnan Yang
✉ lny5400@163.com

RECEIVED 07 April 2024

ACCEPTED 08 July 2024

PUBLISHED 22 July 2024

CITATION

Dang D, Zhang L, Gao L, Peng L, Chen J and
Yang L (2024) Analysis of genomic copy
number variations through whole-genome
scan in Yunling cattle.
Front. Vet. Sci. 11:1413504.
doi: 10.3389/fvets.2024.1413504

COPYRIGHT

© 2024 Dang, Zhang, Gao, Peng, 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.

Analysis of genomic copy number variations through whole-genome scan in Yunling cattle

Dong Dang^{1,2,3}, Lilian Zhang^{1,2,3}, Lutao Gao^{1,2,3}, Lin Peng^{1,2,3},
Jian Chen^{1,2,3} and Linnan Yang^{1,2,3*}

¹College of Big Data, Yunnan Agricultural University, Kunming, China, ²Yunnan Engineering Technology Research Center of Agricultural Big Data, Kunming, China, ³Yunnan Engineering Research Center for Big Data Intelligent Information Processing of Green Agricultural Products, Kunming, China

Yunling cattle is a new breed of beef cattle bred in Yunnan Province, China, which has the advantages of fast growth, excellent meat quality, improved tolerance ability, and important landscape value. Copy number variation (CNV) is a significant source of gene structural variation and plays a crucial role in evolution and phenotypic diversity. Based on the latest reference genome ARS-UCD2.0, this study analyzed the genome-wide distribution of CNVs in Yunling cattle using short-read whole-genome sequencing data ($n = 129$) and single-molecule long-read sequencing data ($n = 1$), and a total of 16,507 CNVs were detected. After merging CNVs with overlapping genomic positions, 3,728 CNV regions (CNVRs) were obtained, accounting for 0.61% of the reference genome. The functional analysis indicated significant enrichment of CNVRs in 96 GO terms and 57 KEGG pathways, primarily related to cell adhesion, signal transduction, neuromodulation, and nutritional metabolism. Additionally, 111 CNVRs overlapped with 76 quantitative trait loci (QTLs), including Subcutaneous fat thickness QTL, Longissimus muscle area QTL, and Marbling score QTL. Several CNVR-overlapping genes, including *BZW1*, *AOX1*, and *LOC100138449*, overlap with regions associated with meat color and quality QTLs. Furthermore, Vst analysis showed that *PSMB4*, *ERICH1*, *SMC2*, and *PPP4R3A* were highly divergent between Yunling and Brahman cattle. In summary, we have constructed the genomic CNV map of Yunling cattle for the first time using whole-genome resequencing. This provides valuable genetic variation resources for the study of the Yunling cattle genome and contributes to the study of economic traits in Yunling cattle.

KEYWORDS

Yunling cattle, copy number variation, single-molecule sequencing, Vst, genome selection

1 Introduction

The cattle (*Bos taurus*), since its domestication about 10,000 years ago, has been a multipurpose domestic animal, occupying an important position in the development of animal husbandry and holding significant importance in national economic growth. It provides human with livestock products such as milk, meat, and leather, and is also used for cultivation and transportation (1). Environmental factors, geographic isolation, and human activities have significantly impacted the genome of cattle, contributing to the development of modern cattle and determining their phenotype, adaptability, and productive performance. Genetic variants, such as single nucleotide polymorphisms (SNPs) and insertion deletions (INDELs),

have been extensively studied in different cattle populations to understand cattle evolution, including population structure, selection, population history and gene introgression (2–4). Some candidate genes related to reproduction, meat, milk and environmental adaptation have been identified (5).

Copy number variation (CNV) is a significant aspect of genomic structural variation. It refers to duplications or deletions of genomic segments, ranging in size from 50 bp to several Mp, which vary among individuals or species (6). CNV can interfere with gene expression and exert a greater impact on phenotypes compared to SNPs (7, 8). The current methods for detecting CNVs include comparative genomic hybridization arrays (CGH arrays), SNP arrays (such as the Illumina BovineHD BeadChip and Illumina BovineSNP50 BeadChip), whole genome sequencing, and long-reads sequencing. While whole genome sequencing and long-reads sequencing offer higher precision breakpoints, sensitivity, and resolution compared to array technologies, limited studies have been conducted to detect CNVs in cattle genomes using long-reads sequencing due to its high cost (5).

The Yunling cattle is the fourth new breed of beef cattle with completely independent intellectual property rights bred by the Yunnan Academy of Grassland and Animal Science, and it is the first breed bred in China through the ternary crossbreeding method. Yunling cattle are a beef cattle breed that has been carefully cultivated for over 30 years. They are selected from three breeds of cattle: Brahman cattle, Murray Grey cattle, and Yunnan Yellow cattle, using crossbreeding selection. It is characterized by rapid growth, high reproductive survival rates, good heat tolerance, and excellent meat quality (9). To date, only a limited number of studies have investigated the genomic distinctions between Yunling cattle and other breeds of cattle at the level of SNPs and INDELs. These differences have identified candidate genes associated with growth, muscle development, neurotransmitter concentration, and heat tolerance (9, 10). However, the research on copy number variation in Yunling cattle only focuses on the impact of CNV of a single gene on growth traits, such as *VAMP7*, *DYNC1I2*, *PLA2G2A*, *SYT11*, etc. (11–14).

This study combines high-coverage short-read data and long-read data for genome-wide CNV analysis, with a view to generate a comprehensive CNV landscape of Yunling cattle, investigating and compare the diversity and population genetic characteristics of CNV regions (CNVRs) in Yunling cattle. In addition, we conducted in-depth analyses of CNV functions and explored the population genetic characteristics of CNV using selective sweep analysis. This laid the foundation for determining the formation mechanism of economically important traits in Yunling cattle and provided a theoretical basis for future Yunling cattle breeding.

2 Materials and methods

2.1 Samples collection and genome sequencing

In this study, we collected sequencing datasets of 130 Yunling cattle and 10 Brahman cattle (Supplementary Table 1). The short-sequencing datasets of 129 Yunnan Ling cattle and 10 Brahman cattle were downloaded from the NCBI public database under

the BioProject accession number PRJNA555741. Another sample data was obtained from a 4-year-old male Yunling cattle reared at Chuxiong Jinda Farm, Chuxiong City, Yunnan Province. Genomic DNA was extracted from heart tissue using the standard phenol-chloroform extraction method for DNA sequencing library construction. The integrity of genomic DNA molecules was checked using agarose gel electrophoresis. The BGISEQ DNBSEQ-T7 platform was used for short sequencing (bp) to obtain 161.89 GB of raw data (64X coverage of the estimated genome size), and the PacBio Sequel II platform (CCS mode) was used for single-molecule long-read sequencing to obtain 61.81 GB of raw data for genome assembly. The sequencing work has performed at GrandOmics Biosciences Co., Ltd. (Wuhan, China).

2.2 Sequence alignment to reference genome

After obtaining the downloaded and sequenced raw data from the whole genome sequencing of Yunling cattle, fastp 0.23.4 (<https://github.com/OpenGene/fastp>) was used to filter the raw data for quality control and retain the relatively high quality sequencing data. The filtered data were then aligned to the latest cattle reference genome ARS-UCD2.0 downloaded from the Ensembl website using BWA-mem with default settings, and the PCR duplicates that could affect the CNV analysis were removed using Picard 2.9.2 (<https://broadinstitute.github.io/picard/>) and Markduplicates. Additionally, PacBio long-read were mapped to the cattle genome (ARS-UCD2.0) using minimap2 (15) with default settings.

2.3 Detection of CNVs and CNVRs

We used Lumpy and CNVcaller to identify CNVs in Illumina short-read sequencing data, and Sniffles (6) to identify CNVs in PacBio long-read sequencing data. We then merged the CNVcaller results with the Sniffles and lumpy results, aiming to maximize the retention of population-specific variants while reducing rare variants at the individual level. Finally, we obtained results based on CNVcaller corrected for individual PacBio data. Specific details are as follows: Sniffles (version: 1.0.10) was employed to detect structural variants (SVs) based on PacBio long reads with default parameters. The SV analysis outputs were filtered through three steps: (1) removed ambiguous breakpoints (flag: IMPRECISE) and low quality SVs; (2) removed SVs shorter than 50 bp; (3) SVs with less than four supporting reads were removed[8]. CNVs were detected using Lumpy software (v 0.2.13) with default parameters, and CNVs were generated by performing discordant-read pairs and split-read pairs on each sample through the lumpyexpress module. CNVcaller was then utilized to detect CNVs. Manual checking and SURVIVOR (version 0.0.1) (16) were used to combine the results of the three software to determine the final dataset. The CNVs of Yunling cattle is divided into duplication CNVR, deletion CNVR and duplication-deletion CNVR, and the length of CNVR is ≤ 50 kb (deletion and both), and the length of CNVR is < 500 kb (duplication) (17). In addition, the

distribution of these regions on the chromosomes of Yunling cattle was analyzed using Bioconductor's RIdeogram software package (18).

2.4 Functional annotation and enrichment analysis of CNVRs

To elucidate the functions associated with the identified CNVs in the Yunling cattle genome, the Yunling cattle annotation file used in this study was downloaded from the NCBI database as ARS-UCD2.0, and the annotation of candidate CNVRs was performed using ANNOVAR (19). GO enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/>) (20) for protein-coding genes only. Biological processes, cellular components and molecular functions were used as GO term categories with a significance level of p -value of 0.05. Additionally, cattle quantitative trait loci (QTL) were downloaded from the cattle QTLdb (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/summary>) (21) and compared with the identified CNVRs. As there are no studies of relevant QTL by ARS-UCD2.0, we only considered QTL reported in ARS-UCD1.2 with confidence intervals < 5 Mb. We used the Bedtools-v2.27.1 (22) "intersect" command to detect which QTL overlapped with identified CNVRs.

2.5 Sweep selective analysis of the CNVR

We calculated the V_{st} (23) between Yunling cattle and Brahman cattle to identify the highly differentiated regions between the two populations. Due to the large difference in the amount of data between the two populations, we randomly selected ten of the 130 Yunling cattle short-read data and one Yunling cattle long-read data to form the Yunling cattle dataset for V_{st} analysis. V_{st} is a method similar to the F_{st} method for calculating selection between populations and is therefore used to calculate data on population differences based on copy number. The formula is $V_{st} = (V_t - V_s)/V_t$, where V_t is the variance between all uncorrelated individuals and V_s is the average variance within each population, weighted according to population size (24). Subsequently, CNVRs with the top 10% of V_{st} values were then used as candidate CNVRs and functional enrichment analysis was performed on these regions.

3 Results

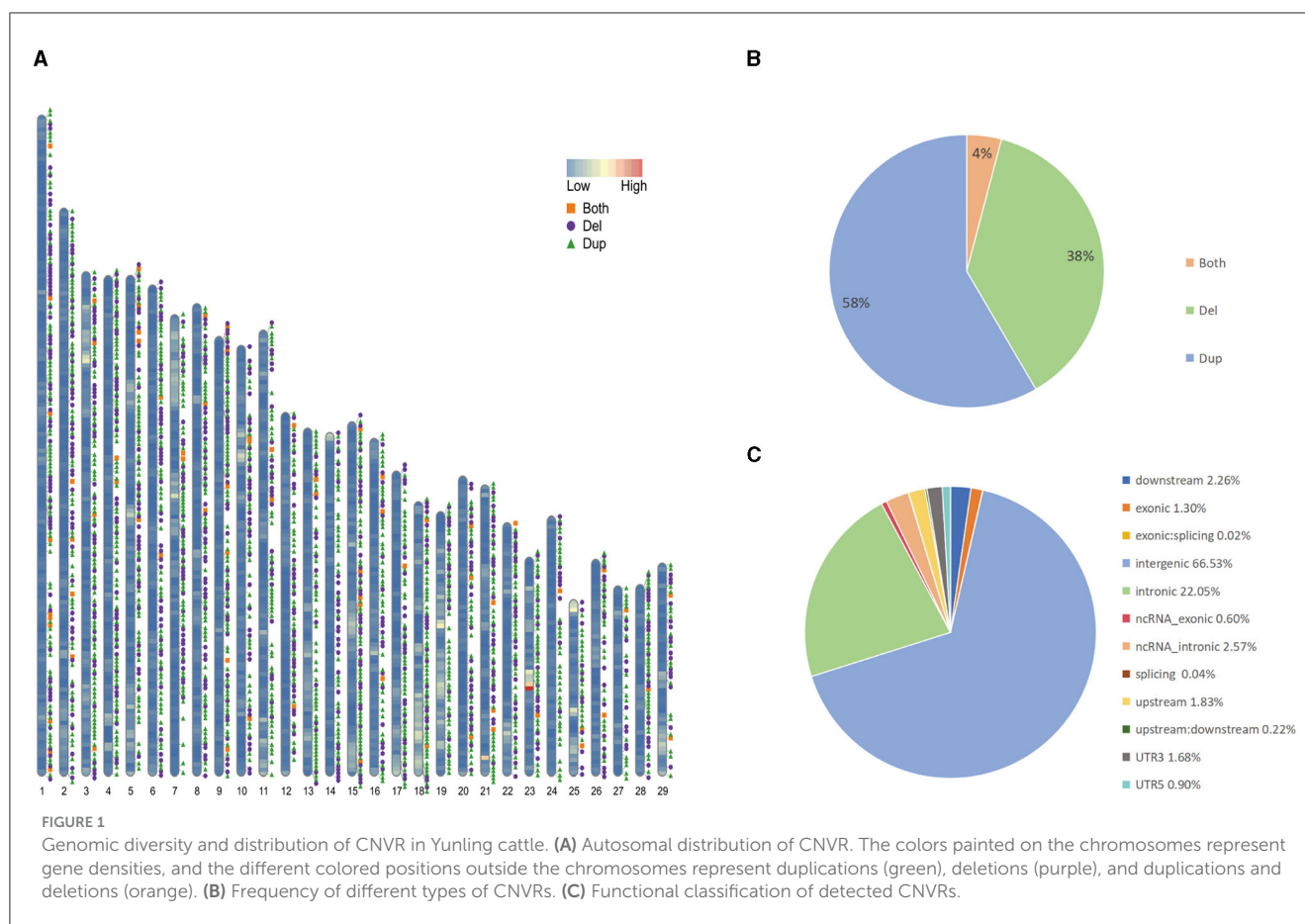
3.1 The landscape of copy number variation in Yunling cattle

We collected sample data from 130 Yunling cattle and 10 Brahman cattle. Among the 140 samples, we newly performed long-read and short-read sequencing on 1 Yunling cattle sample and obtained 61.81 and 161.89 GB of raw data, respectively. The other 139 genome sequences are available online. Among them, the average coverage depths of 25 X and 64 X were obtained

for the self-sequenced long and short data, and 4.2263 X to 7.529 X for the online-accessible data, and an average coverage of 99.73% was obtained after mapping the reads to the cattle reference genome ARS-UCD2.0 (Supplementary Table 1). An average coverage depth between 4x and 8x, as reported in the literature, can provide sufficient power for CNV detection using read depth-based methods (25). We constructed confidential CNV datasets by applying three software packages, Sniffles, lumpy and CNVcaller, to Nanopore long-read sequencing data and Illumina short-read sequencing data. We generated CNVR datasets for two cattle breeds, with a total of 16,507 CNVs detected in Yunling cattle. After merging CNVs with overlapping genomic locations, 3,728 CNV regions (CNVRs) were obtained (Figure 1A; Supplementary Table 2), including 2,175 duplication CNVRs, 1,401 deletion CNVRs, and 152 duplication and deletion CNVRs (Figure 1B), and a total of 3,633 CNVRs detected in Brahman cattle, including 2,186 duplication CNVRs; 1,321 deletion CNVRs; and 126 duplication and deletion CNVRs. Here, we focused on analyzing the CNVRs of Yunling cattle with a total length of 16,392,409 bp and an average length of 4,397 bp, covering 0.61% of the reference genome. The length of CNVRs was mainly distributed in the range of 1,000–2,000 bp, accounting for about 76% of the detected CNVRs (Supplementary Figure 1). Furthermore, the number of CNVRs decreased with increasing length (>500 bp). Furthermore, our results showed that CNVRs were unevenly distributed throughout the genome, with 66.5% of CNVRs (3,676) located in intergenic regions and only 1.3% in exonic regions (Figure 1C).

3.2 Functional annotation of CNVRs

Genes annotated corresponding to 3,728 CNVRs using the NCBI ARS-UCD2.0 reference genome. When CNVRs and genes overlapped by more than 1 bp, the relevant gene was annotated, otherwise the closest gene was annotated. When other genes were present within 1kb, additional annotation was required because it could affect the expression of that gene. Eventually, 3,728 CNVRs were annotated to a total of 1,572 genes. In order to further understand the effects of CNV on various aspects of growth in Yunling cattle, this study used the DAVID website to perform GO enrichment analysis and KEGG pathway analysis on the genes in the CNVRs. The result showed 96 GO terms were enriched (p -value >0.01), including 35 biological processes, 36 cellular components and 25 molecular functions (Supplementary Table 3), which were involved in cell adhesion, neuromodulation, immunomodulation and metabolism. Specifically, ATP binding (GO:0005524), calcium ion binding (GO:0005509), signal transduction (GO:0007165), neuron projection (GO:0043005), actin binding (GO:0003779), etc. In addition, KEGG pathway analysis showed that CNVR-carrying genes were enriched in 57 pathways (p -value >0.05, Figure 2; Supplementary Table 4), including signal transduction (bta04014:Ras signaling pathway, bta04015:Rap1 signaling pathway, bta04022:cGMP- PKG signaling pathway, bta04024:cAMP signaling pathway and bta04020:Calcium signaling pathway), nutrient metabolism (bta00230:Purine metabolism, bta04911. Insulin secretion and bta04974:Protein



digestion and absorption), Regulation of lipolysis in adipocytes (bta04923) and ABC transporters (bta02010).

3.3 QTLs overlapping with identified CNVRs

In order to further reveal the CNVRs associated with economic traits in Yunling cattle and to verify their genetic effects, the detected CNVRs were compared with QTLs in the cattle QTL database. The results showed that 111 CNVRs overlapped with 76 QTLs by more than 1 kb, including Subcutaneous fat thickness QTL (17 CNVRs), Longissimus muscle area QTL (16 CNVRs), Multiple birth QTL (12 CNVRs), Marbling score QTL (8 CNVR), meat color and quality QTL (10 CNVRs) (Supplementary Table 5). Some CNVR genes associated with economic performance of Yunling cattle were identified, such as *XKR4*, *ZBTB7C*, *CCDC15* genes relevant to Longissimus muscle area QTL. *NHLRC3*, *LOC132342211*, *SLCO3A1* genes relevant to Marbling score QTL. *BZW1*, *AOX1*, *LOC100138449* genes linked with meat quality QTL.

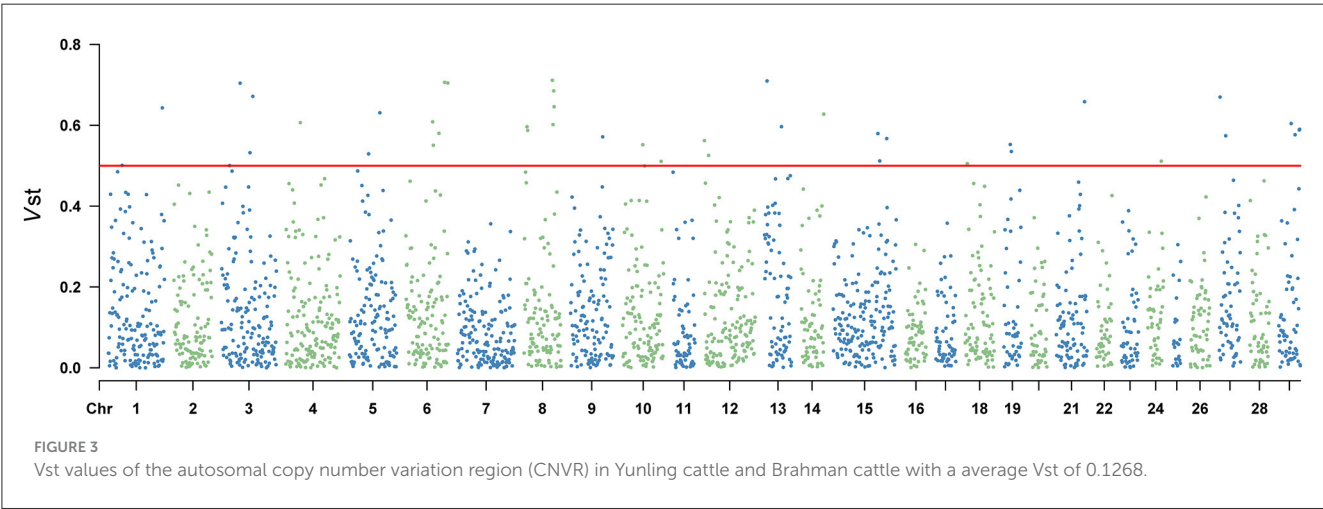
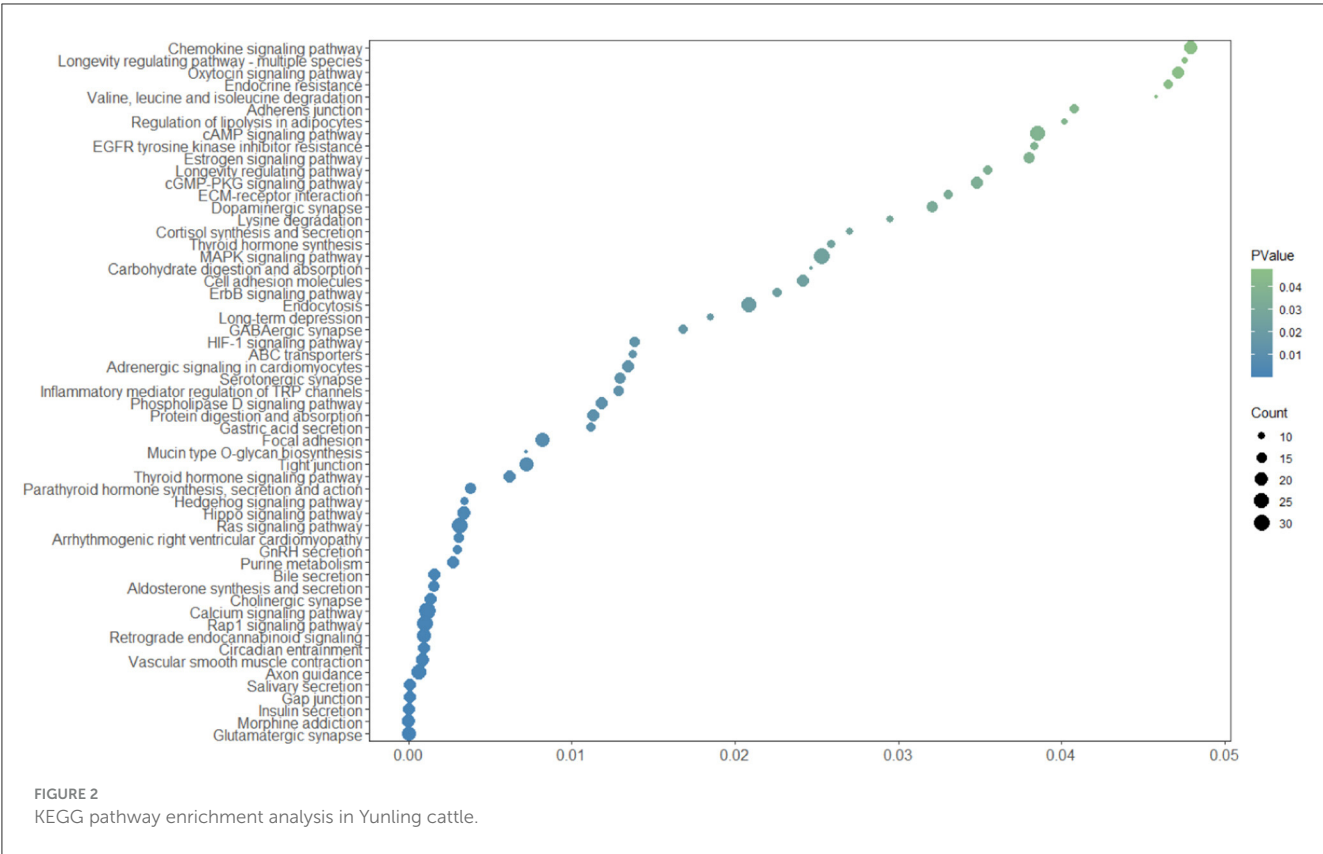
3.4 CNVRs diverging among populations

Vst statistic was utilized to analyze the population differentiation of CNVR between Yunling cattle and Brahman

cattle. The method was similar to Fst in estimating population-specific selection pressure at the gene level. However, Vst showed the average value of detected response CNVR reached 0.3061 by utilizing CNVR-annotated protein-coding genes (Supplementary Table 6). As shown in Figure 3 and Supplementary Table 6, the different CNVRs were unevenly distributed across the chromosomes, and the Manhattan plot showed the results of Vst with the chromosomes in the horizontal coordinates and the VST values in the vertical coordinates. To understand the genes with a high degree of variation between varieties, genes with Vst > 0.5 (up to 98th percentile) were examined. Result revealed that there were 10 CNVR overlapping genes or loci, including *LOC101906606*, *LOC132346850*, *LOC112447126*, *POLN*, *LOC615258*, *SMC2*, *PSMB4*, *ERICH1*, *PPP4R3A*, *LARGE1*. Remarkably, *PSMB4*, *ERICH1*, *SMC2*, and *PPP4R3A* genes play crucial roles in growth and development.

4 Discussion

During domestication and diversification, the frequency of CNV in the genomes of species responds to selection pressures. Considerable effort has been devoted to identifying causal mutations and genes. However, screening selected genomic copy number genetic markers is complex. Although population genetics in cattle has been extensively studied based on SNPs, little is known about the effects of CNVs on phenotypic and evolutionary



traits. CNVs cover a larger region of the genome than SNPs and can affect gene function in a number of ways, including altering gene structure and dosage, altering gene regulation and exposing recessive alleles (26). Over the past decades, high-throughput sequencing technologies and bioinformatics tools have been increasingly used to construct genome-wide CNV profiles (17). CNV diversity has been extensively explored in *Bos Taurus*, *Bos Indicus*, and their crossing populations (27).

Yunling cattle have good fattening performance, significant body proportions, high meat yield, good carcass traits and good fatty acid composition in meat, which is an important source of beef production in China (28). In our study, we firstly used

different sequencing platforms (short-reads and long-reads) to sequence high quality whole genome data of 140 Yunling and Brahman cattle to detect CNV with a high pairwise ratio (average pairwise ratio: 99.73%) compared to the newly reported reference genome (ARS-UCD2.0) (Supplementary Table 1). compared to the UMD 3.1 reference genome. Its improved reliability in screening CNVs (29). Short-reads sequencing has high base identification accuracy, which is an advantage in detecting shorter structural variants, but has an inherent disadvantage of exhibiting a high false-positive rate when detecting complex or long CNVs (30). Long-reads sequencing has the potential to substantially improve the reliability and resolution of structural variant detection. With

an average read length of 10 kbp or more, reads can be more confidently compared to repetitive sequences that often mediate the formation of structural variants. Long-reads sequencing are also more likely to cross structural variant breakpoints with high confidence comparisons. However, long-reads sequencing also have the disadvantage of a high error rate in sequencing (31). Therefore, it happens that long and short reads data complement each other perfectly, providing us with the accuracy of detecting CNVs. We used three software programs for CNV detection, which used different algorithms to obtain accurate estimates of copy number at breakpoints and structurally variable sites, to construct a confidential CNV dataset for Yunling cattle, which also ensured that we obtained a highly confidential CNV dataset. In total, we detected 3,728 CNVRs in 130 Yunling cattle (Figure 1A) and 3,633 CNVRs in 10 Brahman cattle, which are similar to the CNVR levels in other cattle breeds (32). For better statistics, the variants were classified into three categories: duplication, deletion and duplication-deletion. The number of duplication was higher than the number of deletions (Supplementary Table 2). And the length of most CNVRs ranged from 1 to 2 kb (Supplementary Figure 1). In addition, the locations of CNVRs were not uniformly distributed in the cattle genome and they were not randomly distributed on chromosomes. Annotation revealed that CNVRs were mostly annotated in intergenic or intronic regions of the cattle genome (Figure 1C). Previous studies also support that many CNVRs are located on highly variable genes (17).

In this study, the ANNOVAR tool was used to identify genes located within CNVRs, and then DAVID was used to search for GO terms and KEGG pathway information for genes contained within CNVRs, and the purpose of obtaining these results was to speculate on the functions of these genes and the possible effects of CNVRs on economic traits in Yunling cattle. Existing studies have found that genes in CNVRs are mainly involved in immunity, sensory perception of the external environment (involving olfaction, vision, and taste), response to stimuli, and neurodevelopment, with relatively little involvement with nucleic acid binding, metabolism, and cell proliferation (33). This may be due to the fact that the emergence of genes with important functions CNV is the result of strong purifying selection. When CNVs occur in coding regions where the affected gene plays an important role in growth, it is highly likely that the occurrence of the variant will be harmful to the organism and therefore quickly eliminated. This is consistent with the results of our analyses. Three thousand seven hundred and twenty-eight CNVRs were annotated to a total of 1,572 genes. GO analyses revealed that these genes were mainly cell adhesion, neuromodulation, immunomodulation, and metabolism, while a small number of genes were involved in cell differentiation and organ development (Supplementary Table 3). Independent CNVs between different breeds may contribute to breed differences to varietal differences (34). In KEGG pathway analysis, CNVR genes were significantly enriched in signal transduction, nutrient metabolism, and ABC transporter proteins (Figure 2; Supplementary Table 4). Studies in mammals have revealed that ABC transporter proteins can carry a variety of endogenous metabolites, such as amino acids, peptides, and sugars, across lipid membranes, thereby facilitating the absorption and utilization of these nutrients (35).

The CNVRs detected in this study were compared with the QTLs reported in the cattle QTL database. It was found that QTL affecting economic traits of livestock genetic variation can be identified by screening the genome for relevant genes contained in the CNVR (36). After integrating CNVR into QTL, we identified 111 CNVR overlapping with 76 cattle QTL regions in this study (Supplementary Table 5). Many of the CNV overlapping genes, such as *XKR4*, *ZBTB7C*, *CCDC15*, *AOX1*, were located in growth and carcass QTL regions. The *XKR4* gene was identified to be associated with growth traits especially Heritabilities for carcass weight (37, 38). The *ZBTB7C* gene controlled the expression of *MMPS*, which is the zinc-containing endopeptidases that play roles in cell proliferation, migration, differentiation, angiogenesis, and apoptosis (39). *CCDC15*'s role in recruiting both the inner scaffold protein *POC1B* and the distal *SFI1*/Centrin-2 complex to centrioles (40). The levels of 2-pyrrolidone and glycerophospholipids are regulated by the gene expression of *AOX1*, which further affects the levels of volatiles, 2-pyrrolidone and decanal, respectively (41). *NHLRC3*, *LOC132342211*, and *SLCO3A1*, are located in the marbling scoring QTL region. Therefore, these identified CNV-carrying genes provide candidate molecularly relevant markers for future Yunling cattle breeding.

Selective scanning can reveal putative regions subject to environmental and artificial selection during local adaptation and domestication (9). In genomes, calculation of paired Vst values can be used to screen for key CNVRs that differ significantly between populations (42). In this study, five CNVRs carrying the *PSMB4*, *POLN*, *LARGE1*, *SMC2* and *PPP4R3A* genes showed significant pairwise differentiation between Yunling Cattle and Brahman cattle (Figure 3; Supplementary Table 6). *PSMB4* has been reported as a key gene regulating muscle growth and development, which determines postnatal growth rate, muscle fiber diameter and density, and fiber type in pigs (43). This function of *PSMB4* gene may play an important role in good carcass traits in Yunling cattle. Subsequently, Yang et al. also demonstrated that *PSMB4* overexpression inhibited cardiomyocyte apoptosis and *I κ B α* expression, promoted the activation of *NF- κ B* (44). DNA polymerase γ (pol γ), encoded by the *PLON* gene, is an A-family DNA polymerase in vertebrates and some other animal lineages. Takata et al. showed that a function of pol γ in meiotic homologous recombination in processing specific substrates (45). A previous study found that Large myd/Largemyd (myd) mice lack expression of like-acetylglucosaminyltransferase-1 (*Large1*) and exhibit severe muscle pathophysiology, impaired mobility, and a markedly reduced life span (46). The *SMC2* gene is considered a candidate gene associated with growth and meat production traits in sheep (47). The Protein phosphatase 4 regulatory subunit 3A (*PPP4R3A*), forms a highly conserved trimeric complex called protein phosphatase 4 (*PP4*) with *PPP4C* and *PPP4R2*. This complex plays a role in regulating the cellular processes, including DNA damage repair. *PPP4R3A* is widely expressed in various tissues and organs, participating in multiple cellular functions, such as cell proliferation, apoptosis, and cell cycle regulation (48). Some of the genes associated with physical traits in Yunling cattle were artificially selected in a targeted manner during domestication. Thus, under these selective pressures, CNVs may accumulate in Yunling cattle

populations, thus forming the genetic basis for economically important traits.

5 Conclusion

In current study, we conducted comprehensive analyses to explore genetic variation in Yunling cattle. Based on a high-quality cattle reference genome, we constructed a CNV map of Chinese Yunling cattle using whole genome resequencing data. We defined common and breed-specific CNVRs and further analyzed the possible functions of overlapping CNVR genes using enrichment analysis and QTL database search. Based on paired Vst statistics, we examined CNVR-based population differentiation between Yunling cattle and Brahman cattle and revealed potential genomic regions that may be subject to selection. Our results provide a valuable resource for genome-wide variation in Yunling cattle and help to elucidate the genetic basis of superior traits in Yunling cattle. In addition, these results will contribute greatly to the future selection and development of economic traits in Yunling cattle.

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 approved by Yunnan Agricultural University Life Sciences Ethics Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

DD: Writing – original draft. LZ: Data curation, Writing – review & editing. LG: Supervision, Writing – review & editing. LP: Methodology, Writing – review & editing. JC: Validation, Writing – review & editing. LY: Funding acquisition, Resources, Writing – review & editing.

References

1. Magee DA, MacHugh DE, Edwards CJ. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. *Anim. Front.* (2014) 4:7–22. doi: 10.2527/af.2014-0017
2. Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* (2014) 10:e1004254. doi: 10.1371/journal.pgen.1004254
3. Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, et al. The genome landscape of indigenous African cattle. *Genome Biol.* (2017) 18:1–14. doi: 10.1186/s13059-017-1153-y
4. Kim K, Kwon T, Dessie T, Yoo D, Mwai OA, Jang J, et al. The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nat Genet.* (2020) 52:1099–110. doi: 10.1038/s41588-020-0694-2
5. Sun T, Pei S, Liu Y, Hanif Q, Xu H, Chen N, et al. Whole genome sequencing of simmental cattle for SNP and CNV discovery. *BMC Genom.* (2023) 24:179. doi: 10.1186/s12864-023-09248-x
6. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature.* (2011) 470:59–65. doi: 10.1038/nature09708

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Research and Demonstration on Intelligent Management of High Quality Beef Cattle Industry in Yunnan Plateau (Major Special Project of Yunnan Province, No. 202102AE090009).

Acknowledgments

We are grateful to the Yunnan Academy of Grassland and Animal Science for generously providing samples and invaluable support. We also appreciate the computational resources provided by the Yunnan Agricultural Big Data Center.

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.2024.1413504/full#supplementary-material>

7. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet.* (2013) 14:125–38. doi: 10.1038/nrg3373
8. Xu Y, Shi T, Cai H, Zhou Y, Lan X, Zhang C, et al. Associations of MYH3 gene copy number variations with transcriptional expression and growth traits in Chinese cattle. *Gene.* (2014) 535:106–11. doi: 10.1016/j.gene.2013.11.057
9. Chen Q, Zhan J, Shen J, Qu K, Hanif Q, Liu J, et al. Whole-genome resequencing reveals diversity, global and local ancestry proportions in yunling cattle. *J Anim Breed Genet.* (2020) 137:641–50. doi: 10.1111/jbg.12479
10. Jia P, Cai C, Qu K, Chen N, Jia Y, Hanif Q, et al. Four novel SNPs of MYO1A gene associated with heat-tolerance in Chinese cattle. *Animals.* (2019) 9:964. doi: 10.3390/ani9110964
11. Liu X, Yang P, Sun H, Zhang Z, Cai C, Xu J, et al. CNV analysis of VAMP7 gene reveals variation associated with growth traits in Chinese cattle. *Anim Biotechnol.* (2023) 34:1095–101. doi: 10.1080/10495398.2021.2011741
12. Li X, Ding X, Liu L, Yang P, Yao Z, Lei C, et al. Copy number variation of bovine DYNC112 gene is associated with body conformation traits in Chinese beef cattle. *Gene.* (2022) 810:146060. doi: 10.1016/j.gene.2021.146060
13. Yang H, Yue B, Yang Y, Tang J, Yang S, Qi A, et al. Distribution of copy number variation in syt11 gene and its association with growth conformation traits in Chinese cattle. *Biology.* (2022) 11:223. doi: 10.3390/biology11020223
14. Yang P, Cai C, Niu M, Liu X, Wang H, Liang H, et al. Effect of copy number variation of PLA2G2A gene to growth traits in Chinese cattle. *Gene.* (2022) 809:146014. doi: 10.1016/j.gene.2021.146014
15. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191
16. Jeffares DC, Jolly C, Hoti M, Speed D, Rallis C, Dessimoz C, et al. Transient structural variations alter gene expression and quantitative traits in *Schizosaccharomyces pombe*. *bioRxiv [preprint].* (2016) 047266. doi: 10.1101/047266
17. Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, et al. An atlas of CNV maps in cattle, goat and sheep. *Sci China Life Sci.* (2021) 64:1747–64. doi: 10.1007/s11427-020-1850-x
18. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, et al. RIdrogram: drawing SVG graphics to visualize and map genome-wide data on the ideograms. *PeerJ Comp Sci.* (2020) 6:e251. doi: 10.7717/peerj-cs.251
19. Wang K, Li M, Hakonarson H. ANNOVAR functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010) 38:e164. doi: 10.1093/nar/gkq603
20. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* (2009) 4:44–57. doi: 10.1038/nprot.2008.211
21. Hu ZL, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* (2019) 47:D701–10. doi: 10.1093/nar/gky1084
22. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033
23. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* (2006) 444:444–54. doi: 10.1038/nature05329
24. Yang L, Xu L, Zhu B, Niu H, Zhang W, Miao J, et al. Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese cattle. *Sci Rep.* (2017) 7:14299. doi: 10.1038/s41598-017-14768-0
25. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* (2012) 22:778–90. doi: 10.1101/gr.133967.111
26. Zhang F, Gu W, Hurler ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genom Hum Genet.* (2009) 10:451–81. doi: 10.1146/annurev.genom.9.081307.164217
27. Liu Y, Mu Y, Wang W, Ahmed Z, Wei X, Lei C, et al. Analysis of genomic copy number variations through whole-genome scan in Chinese Qaidam cattle. *Front Vet Sci.* (2023) 10:1148070. doi: 10.3389/fvets.2023.1148070
28. Chen J, Zhang L, Gao L, Wei Z, Dang D, Yang L. Population structure and genetic diversity of Yunling cattle determined by whole-genome resequencing. *Genes.* (2023) 14:2141. doi: 10.3390/genes14122141
29. Zhou J, Liu L, Reynolds E, Garrick D, Shi Y. Discovering copy number variation in dual-purpose Xinjiang brown cattle. *Front Genet.* (2022) 12:747431. doi: 10.3389/fgene.2021.747431
30. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell.* (2021) 184:3542–58. doi: 10.1016/j.cell.2021.04.046
31. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* (2018) 15:461–8. doi: 10.1038/s41592-018-0001-7
32. Mei C, Junjvlieke Z, Raza SHA, Wang H, Cheng G, Zhao C, et al. Copy number variation detection in Chinese indigenous cattle by whole genome sequencing. *Genomics.* (2020) 112:831–6. doi: 10.1016/j.ygeno.2019.05.023
33. De Smith A, Walters R, Froguel P, Blakemore A. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res.* (2009) 123:17–26. doi: 10.1159/000184688
34. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Bastiaansen JW, et al. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genom.* (2013) 14:1–13. doi: 10.1186/1471-2164-14-449
35. Shi H, Li T, Su M, Wang H, Li Q, Lang X, et al. Identification of copy number variation in Tibetan sheep using whole genome resequencing reveals evidence of genomic selection. *BMC Genom.* (2023) 24:555. doi: 10.1186/s12864-023-09672-z
36. Hu ZL, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res.* (2022) 50:D956–61. doi: 10.1093/nar/gkab1116
37. Alam MZ, Haque MA, Iqbal A, Lee YM, Ha JJ, Jin S, et al. Genome-wide association study to identify QTL for carcass traits in Korean Hanwoo Cattle. *Animals.* (2023) 13:2737. doi: 10.3390/ani13172737
38. Bejarano DH, Martinez RA, Rocha JF. Genome-wide association study for growth traits in Blanco Orejinegro and Romosinuano cattle. *Trop Anim Health Prod.* (2023) 55:358. doi: 10.1007/s11250-023-03743-9
39. Jeon BN, Yoon JH, Kim MK, Choi WI, Koh DI, Hur B, et al. Zbtb7c is a molecular 'off' and 'on' switch of Mmp gene transcription. *Biochim Biophys Acta.* (2016) 1859:1429–39. doi: 10.1016/j.bbagr.2016.09.004
40. Arslanhan MD, Cengiz-Emek S, Odabasi E, Steib E, Hamel V, Guichard P, et al. CCDC15 localizes to the centriole inner scaffold and controls centriole length and integrity. *J Cell Biol.* (2023) 222:e202305009. doi: 10.1083/jcb.202305009
41. Liu D, Zhang H, Yang Y, Liu T, Guo Z, Fan W, et al. Metabolome-based genome-wide association study of duck meat leads to novel genetic and biochemical insights. *Adv Sci.* (2023) 10:2300148. doi: 10.1002/adv.202300148
42. Yuan C, Lu Z, Guo T, Yue Y, Wang X, Wang T, et al. A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. *BMC Genom.* (2021) 22:1–10. doi: 10.1186/s12864-021-07387-7
43. Li Q, Hao M, Zhu J, Yi L, Cheng W, Xie Y, et al. Comparison of differentially expressed genes in longissimus dorsi muscle of Diannan small ears, Wujin and landrace pigs using RNA-seq. *Front Vet Sci.* (2023) 10:1296208. doi: 10.3389/fvets.2023.1296208
44. Yang C, Yu P, Yang F, He Q, Jiang B, Zheng L, et al. PSMB4 inhibits cardiomyocyte apoptosis via activating NF- κ B signaling pathway during myocardial ischemia/reperfusion injury. *J Mol Histol.* (2021) 52:693–703. doi: 10.1007/s10735-021-09977-x
45. Takata Ki, Reh S, Yousefzadeh MJ, Zelazowski MJ, Bhetawal S, Trono D, et al. Analysis of DNA polymerase ν function in meiotic recombination, immunoglobulin class-switching, and DNA damage tolerance. *PLoS Genet.* (2017) 13:e1006818. doi: 10.1371/journal.pgen.1006818
46. Yonekawa T, Rauckhorst AJ, El-Hattab S, Cuellar MA, Venzke D, Anderson ME, et al. Large1 gene transfer in older myd mice with severe muscular dystrophy restores muscle function and greatly improves survival. *Sci Adv.* (2022) 8:eabn0379. doi: 10.1126/sciadv.abn0379
47. Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, et al. Genome-wide association studies for growth and meat production traits in sheep. *PLoS ONE.* (2013) 8:e66569. doi: 10.1371/journal.pone.0066569
48. Hu Y, Han Z, Guo H, Zhang N, Shen N, Jiang Y, et al. Identification of a novel germline PPP4R3A missense mutation Asp409Asn on familial non-medullary thyroid carcinoma. *Biomedicine.* (2024) 12:244. doi: 10.3390/biomedicine12010244



OPEN ACCESS

EDITED BY

Fei Hao,
Northumbria University, United Kingdom

REVIEWED BY

Yufang Liu,
Hebei University of Engineering, China
Ran Li,
Northwest A&F University, China

*CORRESPONDENCE

Yongbin Liu,
✉ ybliu@imu.edu.cn
Shaoyin Fu,
✉ fushao1234@126.com

RECEIVED 10 October 2024

ACCEPTED 19 November 2024

PUBLISHED 11 December 2024

CITATION

Qi Y, He X, Wang B, Yang C, Da L, Liu B, Zhang W,
Fu S and Liu Y (2024) Selection signature
analysis reveals genes associated with tail
phenotype in sheep.
Front. Genet. 15:1509177.
doi: 10.3389/fgene.2024.1509177

COPYRIGHT

© 2024 Qi, He, Wang, Yang, Da, Liu, Zhang, Fu
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.

Selection signature analysis reveals genes associated with tail phenotype in sheep

Yunxia Qi¹, Xiaolong He², Biao Wang², Chaoyun Yang¹, Lai Da²,
Bin Liu³, Wenguang Zhang⁴, Shaoyin Fu^{2*} and Yongbin Liu^{2,5*}

¹College of Animal Sciences, Xichang University, Xichang, China, ²Animal Husbandry Institute, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, China, ³Inner Mongolia BIONEWS Technology Co., LTD., Hohhot, China, ⁴College of Life Sciences, Inner Mongolia Agricultural University, Hohhot, China, ⁵State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock (R2BGL), College of Life Sciences, Inner Mongolia University, Hohhot, China

Tail type of sheep, which may be affected by many genes with a complex mechanism, is an important economic trait concerned by both raiser and consumers. Here, we employed two sheep breeds with extreme phenotypes - Mongolian sheep (short-fat-tailed) and Bamei Mutton sheep (long-thin-tailed) to analyze the genetic differences at the genomic level and find candidate genes associated with tail phenotype. The results of population structure analysis showed that the LD decay rate of Mongolian sheep was greater than that of Bamei Mutton sheep. When $K = 2$, the two populations were obviously separated with a certain degree of mixing. From 49 sheep individuals, 20,270,930 and 2,479,474 SNPs and Indels were identified, respectively. Selection signals were detected based on F_{ST} , π -Ratio, and XP-EHH. These three methods identified 85 candidate genes, of which *PDGFD*, *GLIS1*, *AR*, and *FGF9* were reported to be associated with tail fat deposition, while *VRTN* associated with tail length in sheep tail phenotype; the others were novel genes that may play important roles in sheep tail phenotype formation. Gene annotation revealed that these candidate genes mainly participate in pathways associated with fat deposition or lipid metabolism. This study provided insight into sheep tail type development and a guide for molecular breeding.

KEYWORDS

sheep, tail type, selection signature analysis, F_{ST} , π -ratio, XP-EHH

1 Introduction

Sheep (*Ovis aries*), as a major source of meat, milk, fiber and leather for mankind, has been domesticated since Mesolithic period, ~11,000 years ago (Chessa et al., 2009; Alberto et al., 2018; Deng et al., 2020). During this long procedure sheep has evolved diverse phenotypes such as coat color, horn, tail type, etc., under natural and artificial selection (Kijas et al., 2009). Of these phenotypes, tail type is an important economic trait concerned by both producers and consumers. According to its length and fat deposition, sheep tail can be divided into five major types: long-fat tail, short-fat tail, long-thin tail, short-thin tail, and fat-rumped tail. It is widely believed that the wild ancestors of sheep were thin-tailed, while the fat-tailed sheep breeds emerged as an adaptive response to harsh and challenging environmental conditions (such as climate fluctuation, drought, and food scarcity) (Atti et al., 2004; Pourlis, 2011; Moradi et al., 2012; Kalds et al., 2021). Fat-tailed sheep could deposit up to 20% of their carcass weight as fat in the tail (Yousefi et al., 2012). The large

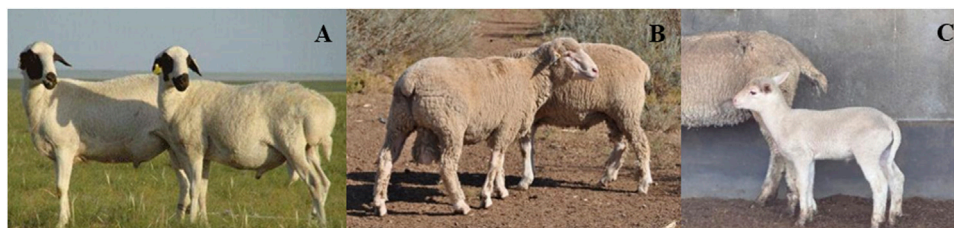


FIGURE 1
Tail phenotype of sheep. (A) Short-fat-tailed Mongolian sheep. (B) Long-thin-tailed Bamei Mutton sheep after docking. (C) long-thin-tailed lamb of Bamei Mutton sheep before docking.

amount of tail fat serves as an energy source for sheep, and also provided people valuable edible fat in the era of material scarcity. Nowadays, however, with the increasing incidences of obesity and cardiovascular disease, people prefer a diet low in fat and high in protein. On the other hand, with the popularization of intensive and semi-intensive feeding management, fat-tail of sheep has lost its original advantages and brought inconvenience to production management, such as inconvenience for mating (Kridli and Said, 1999) and locomotion (Orihuela and Ungerfeld, 2019). In addition, a large amount of fat deposition in tails may reduce feed conversion rate and even affect the carcass quality (Safdarian et al., 2008; Yousefi et al., 2012). Yousefi et al. found that the thin-tailed breed accumulated more intramuscular fat in *longissimus dorsi* muscle and had lower shear force and better eating quality, tenderness, and drip loss than the fat-tailed breed (Yousefi et al., 2012). In practice, tail docking (O'Donovan et al., 1973; Shelton et al., 1991; Bicer et al., 1992; Moharrery, 2007; Wang et al., 2018) and cross-breeding (Kashan et al., 2005; Khaldari et al., 2008; Abdullah et al., 2010) are usually taken to reduce tail size and length. It was reported that tail docking may improve lambs' growth, slaughter performance and mutton quality (Atti and Mahouachi, 2011; Marai et al., 1987; Bicer et al., 1992; Abouheif et al., 1993; Bing et al., 2006). However, tail docking is stressful and risky, and has been banned in several countries to improve animal welfare (Eck et al., 2019). Cross-breeding takes time and efforts, and the results are usually unsatisfactory. Currently, how to breed short-thin-tailed sheep through molecular breeding methods has become a focus of sheep breeders, and the key to solve this problem is to identify genes related to tail phenotype of sheep. There were some research on tail phenotype and several promising genes such as *PDGFD*, *BMP2* and *TBXT*, etc., associated with tail phenotype had been suggested (Yuan et al., 2017; Zhi et al., 2018; Dong et al., 2020; Mastrangelo et al., 2019; Pan et al., 2019; Moradi et al., 2012), but most of the studies focused either on fat deposition or tail length, and the results are usually inconsistent. The genetic mechanics underlying tail phenotype still remain unclear.

Mongolian sheep, short-fat-tailed (Figure 1A), is the most widely distributed and abundant sheep breed in China. It is mainly distributed in Inner Mongolia Autonomous Region, northeast, north, and northwest of China. Mongolian sheep is an ancient indigenous breed formed by natural and artificial selection for a long time, and is favored by local herdsman and consumers because of its rough feeding resistance, cold resistance, drought resistance, and high-quality meat. Bamei Mutton sheep, which is

long-thin-tailed (Figures 1B, C), is the first dual-purpose breed that was bred in China by crossing local fine-mixed sheep as the maternal line with German Merino sheep as the paternal line. It contains 6.25% bloodline of Mongolian sheep. Bamei Mutton sheep is mainly distributed in Bayannur City of Inner Mongolia Autonomous Region, China. It is characterized by resistance to rough forage, strong stress resistance, good adaptability, rapid weight gain in lamb fattening, and early sexual maturity. In the present study, we performed whole genome resequencing of the two breeds with extreme tail phenotypes to investigate selection signatures and candidate genes associated with tail phenotype (fat vs. thin and long vs. short) based on three statistical tests, including fixation index (F_{ST}), π -Ratio, and cross-population extended haplotype homozygosity test (XP-EHH). The candidate genes identified in our study provided the basis for understanding the molecular mechanism of tail phenotype in sheep.

2 Materials and methods

2.1 Sample collection, DNA extraction, and sequencing

A total of 28 Mongolian sheep (MG) and 21 Bamei Mutton sheep (BM) were selected from Inner Mongolia Autonomous Region, China. The sheep were raised and managed under the same condition. All individuals were typical of the breeds and unrelated according to pedigree records or owner's information. Blood samples were collected and returned to laboratory on dry ice. Genomic DNA was extracted from the blood samples following the standard phenol-chloroform extraction procedure. DNA samples that passed the test ($D_{260\text{ nm}}/D_{280\text{ nm}} = 1.7\text{--}1.9$) were randomly interrupted into fragments of 500 bp in length. Paired-end sequencing libraries were constructed according to the manufacturer's instructions (Illumina Inc., San Diego, CA, USA) and sequenced on the Illumina HiSeq Xten Sequencer (Illumina Inc.) with PE150 module.

2.2 Alignments and variant identification

After filtering out low quality reads, the 150-bp paired-end clean reads were mapped onto the sheep reference genome Oar v.4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2) with

BWA-MEM using the default parameters (Li and Durbin, 2009). After alignments, SNP calling was performed using SAMtools and Genome Analysis Toolkit (GATK, v.3.8) (Nekrutenko and Taylor, 2012). All SNPs were filtered using the ‘Variant Filtration’ module of GATK with the standard parameters as below: Window 4; Variants with QD (quality depth) < 4.0; FS (Phred-scaled p -value using Fisher’s exact test to detect strand bias) > 60.0; MQ < 40.0; -G_ filter “GQ < 20”.

The implementation of SAMtools mpileup (v.1.8) (Li, 2011) was run in a multi-sample mode to calculate genotype likelihoods from the aligned reads for all samples simultaneously. The parameters *-E* and *-t* were used to recalculate (and apply) base alignment quality and produce per-sample genotype annotations, respectively. Then, the estimated genotype likelihoods were converted into genotypes using *BCFtools* call using the *-v* and *-m* flags to output variable sites only, and permitted sites to have more than two alternative alleles, respectively.

Based on the annotation file of the sheep reference genome Oar v.4.0, a transcript FASTA file for database was built using the *retrieve_seq_from_fasta.pl* module of ANNOVAR, and then the functional annotation for each SNP was performed using the *table_annoar.pl* module of ANNOVAR (Wang et al., 2010).

2.3 Population structure analysis

SNPs were pruned the in high levels of pair-wise LD using PLINK v.1.9 (Purcell et al., 2007) with the parameter (*--indep-pairwise 50 5 0.2*) to perform principal component analysis (PCA) and ADMIXTURE analysis. PCA of whole-genome SNPs for all 49 individuals was conducted using the GCTA v.1.24.2 (Yang et al., 2011). Furthermore, population structure analysis was carried out using the ADMIXTURE v1.3 (Alexander and Lange, 2011) with kinship (K) ranged from 2 to 5. The unrooted Neighbor-joining (NJ) tree was constructed with TASSEL using the matrix of pairwise genetic distances and visualized with iTOL (<https://itol.embl.de/>). The LD decay for each group was measured using PopLDdecay (Zhang et al., 2019) with default parameters.

2.4 Genome-wide selective sweep test

To identify the selective sweep regions, we performed genome-wide scans of selection signals using three metrics: allele frequency based methods F_{ST} (Weir and Hill, 2002), π -Ratio (Danecek et al., 2011), and haplotype-based method XP-EHH (Sabeti et al., 2007a).

The SNPs were filtered with parameters (*--maf 0.05 --max-missing 0.90*) using PLINK v.1.9 (Purcell et al., 2007). The F_{ST} was calculated using VCFtools (Danecek et al., 2011) with parameter “*--weir-fst-pop group1 --weir-fst-pop group2 --fst-window-size 50000 --fst-window-step 20000 --maf 0.05 --max-missing 0.90*”. Then the F_{ST} values were normalized (ZF_{ST}) using the Ztransformation method (Rubin et al., 2010). The genetic diversity (π -Ratio) was calculated using VCFtools with parameters “*--keep group1/group2 --window-pi 50000 --window-pi-step 20000 --maf 0.05 --max-missing 0.90*” and python scripts. The overlap of the top 5% windows in each method was considered as candidate signatures of selection.

The XP-EHH was performed for every SNP using the default settings by selscan v.1.1 (Szpiech and Hernandez, 2014), and genotypes were phased using Beagle (Browning and Browning, 2007) with default parameters. The genome-wide raw XP-EHH statistics were standardized to a distribution with zero mean and unit variance. SNPs in the top 0.1% are taken as significant SNPs. Significant regions are identified by combining SNPs of significant XP-EHH scores that are less than 200 kb apart. If two SNPs both have significant XP-EHH scores and were less than 200 kb apart, then the two SNPs formed a region.

In the π -Ratio and XP-EHH tests, the BM sheep were used as the target population, and the MG sheep as the reference population.

2.5 Gene ontology enrichment and KEGG pathway analyses

According to genome annotation, a gene was assumed to be under positive selection if it overlapped with a selection signal. To obtain an in-depth view of the biological significance of the candidate genes, online Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted by retrieving *O. aries* in self-built database (No.AH96240) on AnnotationHub website (<https://annotationhub.bioconductor.org/>). Protein-Protein Interaction (PPI) analysis was performed using STRING database (<https://cn.string-db.org/>).

3 Results

3.1 Overview of sequencing quality

After sequencing and data quality control, more than 100 million clean reads were obtained in the MG and BM groups, respectively. The number of clean bases in the BM group was found to be more than 4G greater than that in the MG group. The mapping rate is greater than 98% in both groups with an average depth of $7.31 \times$ (Table 1; Supplementary Table S1).

3.2 SNP identification

In total, 29,926,218 and 23,122,081 SNPs, 3,755,604 and 5,005,409 Indels were identified in MG and BM sheep, respectively, among which, 20,270,930 SNPs and 2,479,474 Indels are common in MG and BM sheep (Figure 2). Functional annotation of the polymorphic sites showed that the vast majority of SNPs and Indels were present either in intergenic regions (62.2% and 60.8%, respectively) or in intronic regions (34.8% and 36.3%, respectively). Exons contained 0.60% of the total variation (Table 2). These results indicate that the variants on the MG and BM genomes differ.

3.3 Population genetic structure

Following the identification of the SNPs, PCA, phylogenetic relationship analysis, and population genetic structure analysis were

TABLE 1 Overview of sequencing statistics.

Group	Clean reads number	Clean bases number (G)	Map rate (%)	Effective depth
MG	113,552,917	17.03	99.30	6.47
BM	143,757,312	21.56	98.73	8.14

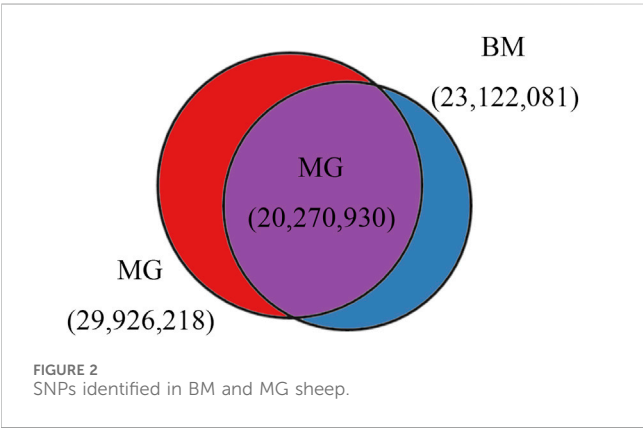


TABLE 2 Annotation of SNPs/Indels.

Variant type	SNP count	Indel count
Intergenic	12,609,201	1,782,495
Intronic	7053401	1062533
Exonic	135627	3,549
Downstream	119577	20132
UTR3	118337	23552
Upstream	111348	18491
UTR5	90819	14489
ncRNA intronic	25381	3,850
Upstream or downstream	3,822	655
ncRNA exonic	2,923	282
Splicing	412	266
Exonic, splicing	16	23

conducted for all the individuals. The PCA results showed that the BM and MG sheep were clearly separated (Figure 3A), and the NJ tree also produced similar results, with BM and MG sheep divided into 2 clades (Figure 3B), which indicated that there is a certain degree of genetic distance between BM and MG sheep. The results of population genetic structure (Figure 3C) showed that for the BM group, the consistency within groups was better when $K = 2$ and $K = 4$, while for the MG, it was better when $K = 2$ and $K = 3$, indicating that the consistency of the individuals within the groups was better, and that differences existed between the groups. Furthermore, when $K = 2$, BM and MG sheep were obviously divided into two subgroups, but with a certain degree of mixing, which was consistent with the fact that BM sheep contains a certain amount

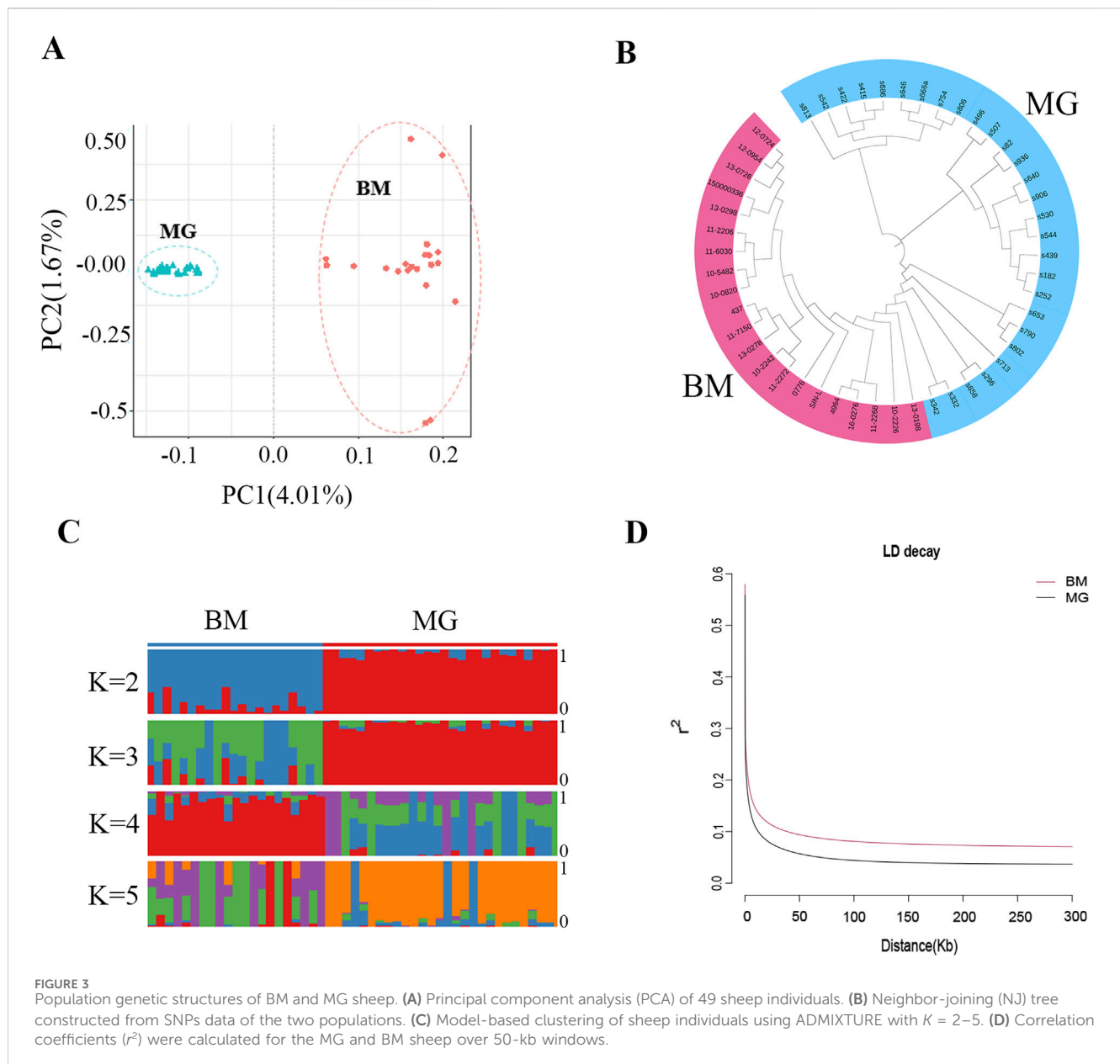
of ancestry of MG sheep. When $K = 3$, no new subgroups appeared in the experimental population. The aforementioned results demonstrate that there are distinctions between BM and MG sheep, which may be attributed to tail phenotype variation. In order to further explore the genetic diversity, linkage disequilibrium, in terms of the correlation coefficient (r^2), was calculated for BM and MG sheep populations. As shown in Figure 3D, the faster LD decay was observed in the MG population, which indicates that MG sheep had higher genetic diversity, and BM sheep had higher degree of domestication and greater intensity of selection.

3.4 Detection of selective sweeps

Due to the genetic separation between MG and BM sheep, selective sweep analysis using F_{ST} , π -Ratio and XP-EHH were performed to investigate selection signals in BM sheep, which would facilitate the identification of target genes. The results demonstrated that the majority of SNPs exhibited moderate genetic variance within the population (Figure 4A, $F_{ST} < 0.15$). Additionally, some SNPs exhibited high genetic variance and high genetic variability on chromosomes 13, 16, and 17, suggesting that these SNPs may be mutations specific to the BM and MG populations. A further 544,123 SNPs were identified under the conditions of $F_{ST} \geq 0.2$ and π -Ratio ≤ 0.397 (Figure 4B). The combined analysis of F_{ST} and π -Ratio revealed that 1884 genes (representing the top 10% of genes) were identified by $\log_2(\pi_{BM}/\pi_{GM})_{ZF_{ST}}$, while 294 genes were identified by XP-EHH (Figure 4C). Eventually, a total of 85 overlapping genes were identified as candidate genes by $\log_2(\pi_{BM}/\pi_{GM})_{ZF_{ST}}$ and XP-EHH (Figure 4D; Supplementary Table S2). Of these candidate genes, some were known to be related to sheep tail phenotype, such as fat deposition associated genes *PDGFD*, *GLIS1*, *AR*, *FGF9*, and vertebral number variation associated gene *VRTN*; some were novel genes that may have relationship with sheep tail phenotype formation.

3.5 Function annotation of the selected genes

Functional association of the 85 candidate genes was further investigated by GO and KEGG analysis. For GO analysis, the biological processes mainly focused on ‘cellular response to organic substance’, ‘cellular response to lipid’, ‘cellular response to organic cyclic compound’, ‘negative regulation of cell development’, ‘beta stimulus regulation of DTPase activity’, and so on (Figure 5A) ($p < 0.05$). In terms of KEGG, several signaling pathways related to lipid metabolism were significantly



enriched, including ‘phospholipase D signaling pathway’, ‘glycerophospholipid metabolism’, ‘glycerolipid metabolism’, and ‘fatty acid elongation’ (Figure 5B). In addition, some significant pathways were also enriched, such as ‘MAPK pathway’, ‘P13K Akt signaling pathway’, ‘Calcium signaling pathway’ and ‘ras signaling pathway’. The interaction of KEGG pathways and the relationship between genes and pathways was shown in Figure 5B. Furthermore, these candidate genes showed close functional association (Figure 6).

4 Discussion

In the past decade, the improvement of sheep tail phenotype has gradually developed from utilization of traditional hybridization or tail-docking to molecular breeding

technology. To date, several methods are employed to detect the selective sweeps in various livestock genomes. In the present study, we used three metrics, allele frequency-based methods F_{ST} , π -Ratio, and haplotype-based method XP-EHH, to identify genome-wide selective sweep regions. The power of each test was different, and any set of candidate genes may contain some false positives (Nielsen et al., 2007). F_{ST} measures the genetic differentiation between populations (Holsinger and Weir, 2009); π -Ratio identifies the differences in nucleotide divergence between populations (Sun et al., 2020); XP-EHH detects ongoing or nearly fixed selective sweeps by comparing haplotypes between the two populations (Sabeti et al., 2007b). Combining multiple tests can improve the power of detecting selection signatures (Zeng et al., 2007), making the results more reliable. We considered the overlapping genes derived from three methods as candidate genes and eventually identified

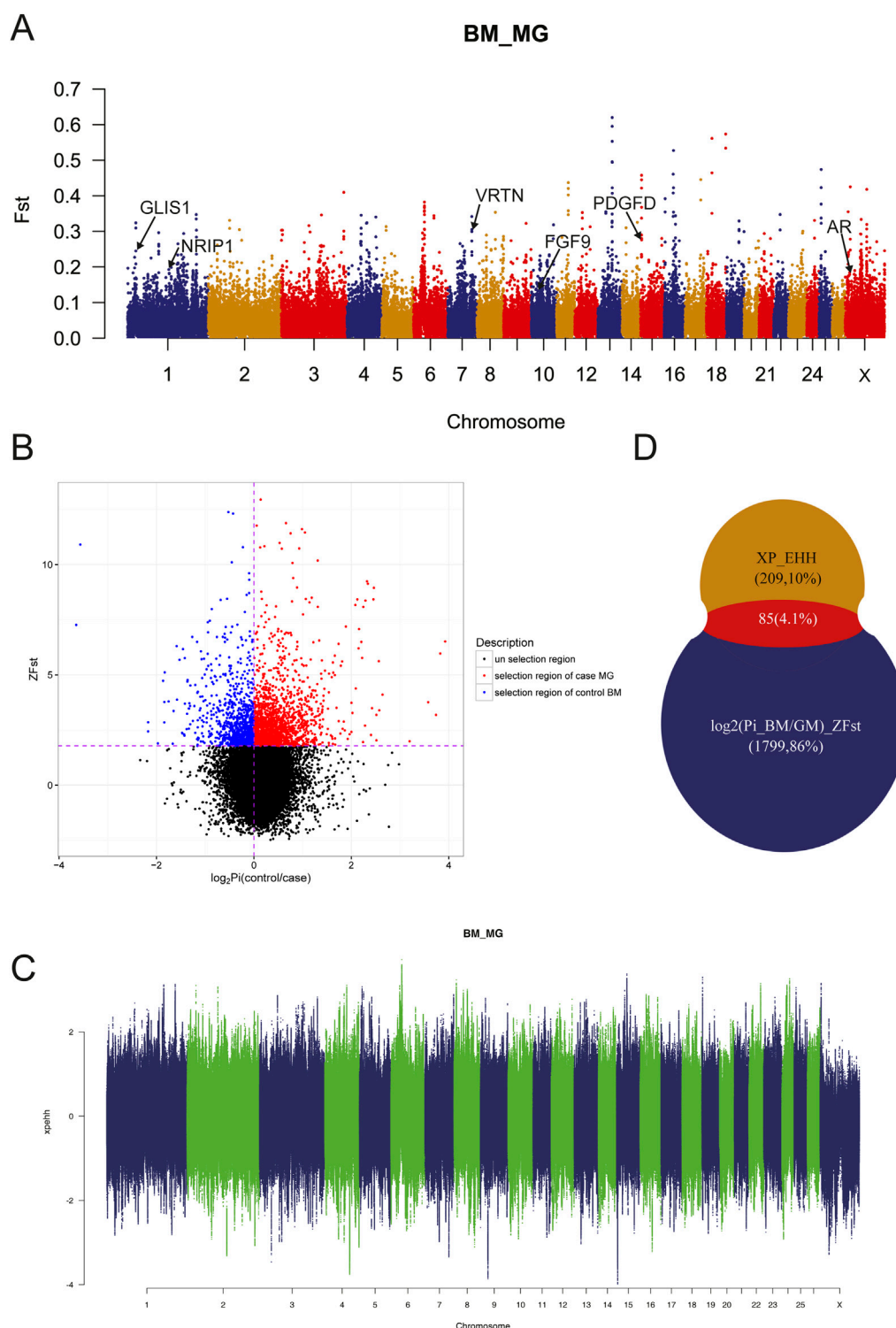
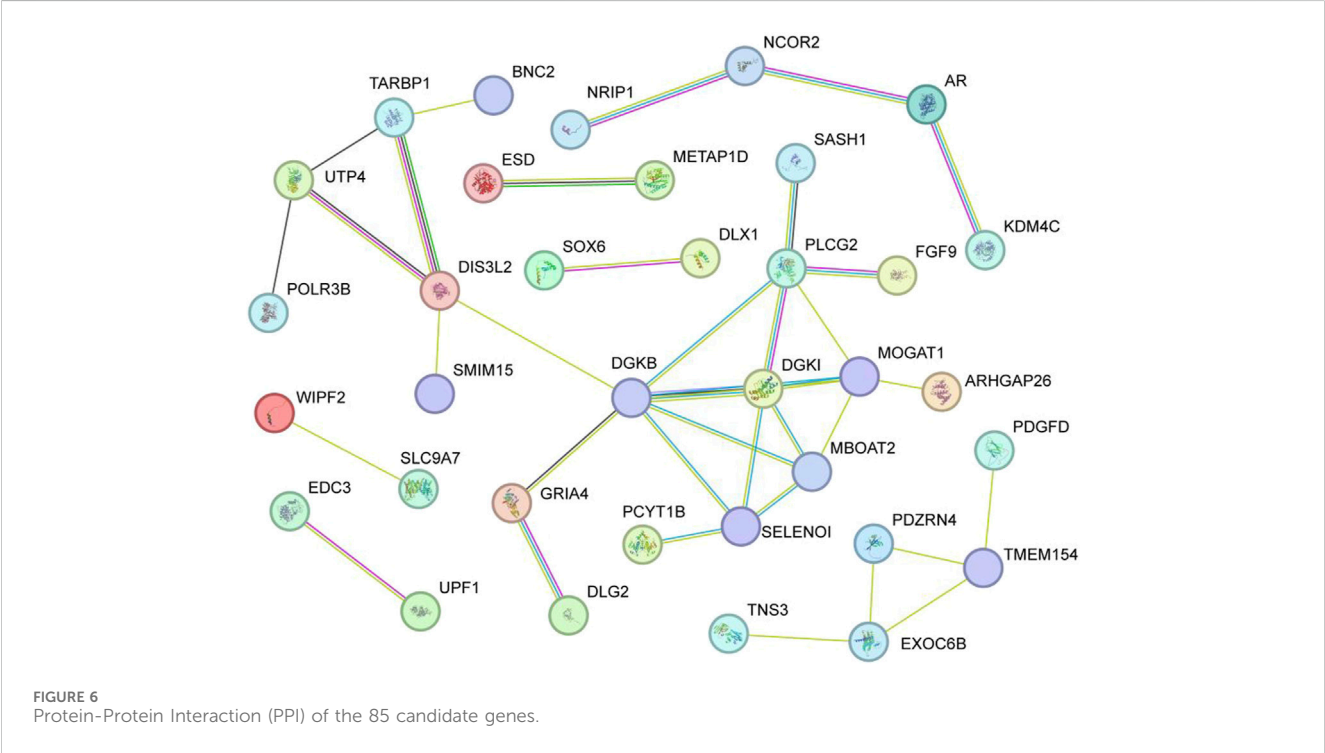
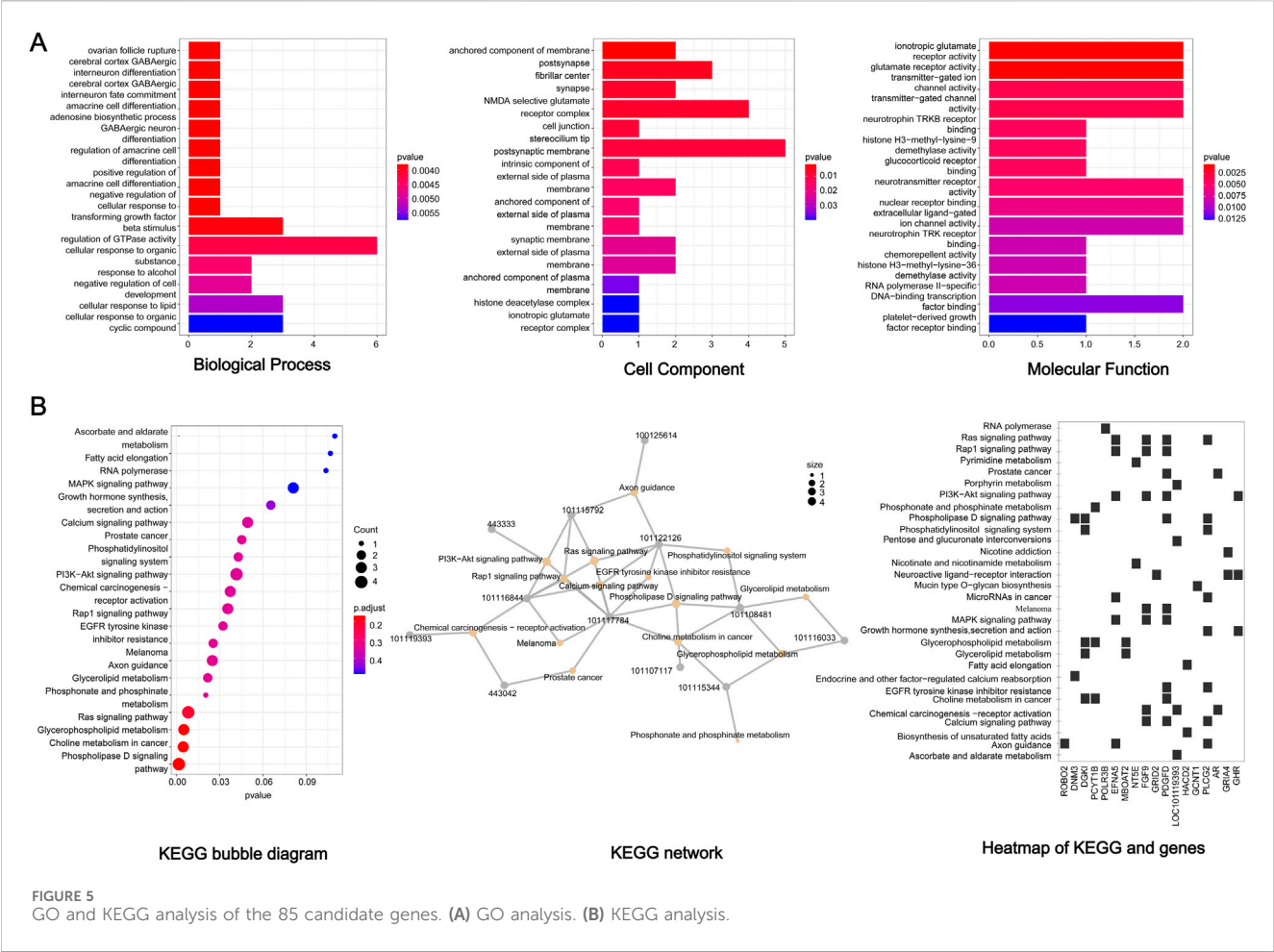


FIGURE 4
Genomic Selection analyses. **(A)** Selective signals detected by F_{ST} . **(B)** Selective signals detected by $\log_2(Pi_BM/GM_ZF_{ST})$. **(C)** Selective signals detected by XP_EHH. **(D)** Overlapping genes identified by $\log_2(Pi_BM/GM_ZF_{ST})$ and XP_EHH.

85 candidate genes. Among these genes, *PDGFD* (Dong et al., 2020; Zhu et al., 2021; Pan et al., 2019; Mastrangelo et al., 2019; Zhao et al., 2020; Li et al., 2020a; Li et al., 2020b; Wang et al., 2022; Wei et al., 2015; Luo et al., 2021), *GLIS1* (Luo et al., 2021), *NRIP1*

(*RIP140*) (Xu et al., 2017), *AR* (Moradi et al., 2022), *FGF9* (Moioli et al., 2015), and *VRTN* (Mastrangelo et al., 2019; Zhu et al., 2021; Moioli et al., 2015) were formerly reported to be involved in regulation of tail fat deposition or tail length.



Many studies have recently highlighted the *platelet-derived growth factor D* (*PDGFD*) gene as a new sheep tail phenotype pattern maker (Wei et al., 2015; Pan et al., 2019; Dong et al., 2020; Li Q. et al., 2020; Li X. et al., 2020; Luo et al., 2021; Mastrangelo et al., 2019). Dong et al. found that the expression of the *PDGFD* gene is higher in fat-tailed breeds than in thin-tailed breeds, and a similar result was observed in obese mice and human after analyzing a public transcriptomic data (Dong et al., 2020). Overexpression of *PDGFD* in ovine preadipocytes could promote adipogenic differentiation, and the expression levels of two adipogenesis marker genes (*PPARc* and *LPL*) increased after *PDGFD* overexpression (Li Q. et al., 2020). Furthermore, oil red O staining showed that the number of lipid drops was higher in the *PDGFD*-overexpressing group than in the control group (Li Q. et al., 2020). These studies indicated that *PDGFD* gene plays a positive regulation role in the fat deposition process of sheep tail. Interestingly, there were also different discoveries about the expression profile of the *PDGFD* gene. Li X. et al. (2020) identified four *PDGFD* transcripts (I, II, III, and IV), and transcript I was the most differentially expressed transcripts between the thin-tailed and the fat-tailed/fat-rumped sheep breeds. Notably, *PDGFD* expression (at the mRNA and protein levels) was consistently negatively correlated with fat deposition in sheep tails (Li X. et al., 2020). The highest *PDGFD* gene expression level was observed in the thin-tailed Chinese Merino sheep, followed sequentially by the small fat-tailed Han sheep, the large fat-tailed Han sheep, and the fat-rumped Altay sheep (Li X. et al., 2020). The authors inclined to the idea of involving the *PDGFRb* signaling (a receptor of *PDGFD*) in inhibiting the differentiation of white adipocytes by regulating the expression of two key transcriptional regulators of adipogenesis (*PPARc2* and *C/EBPα*) (Olson and Soriano, 2011; Kalds et al., 2021). In the present study we found that *PDGFD* gene was strongly selected by Bamei Mutton sheep, combined with previous researches, indicating that *PDGFD* was involved in regulating the fat deposition process of sheep tail, but how to regulate this process still needs more in-depth research.

GLIS1 is a zinc finger protein that acts as both an activator and repressor of transcription (Kim et al., 2002). The temporal and spatial expression of *GLIS1* is consistent with mesoderm differentiation during mouse embryonic development (Nakashima et al., 2002). Later, Tosic et al. discovered that *GLIS1* was highly expressed in bipotent muscle satellite cells. But when overexpressed, increased occupancy of *GLIS1* is observed at the promoters of adipogenic genes *Adipoq*, *Cebpa* and *Ucp1*, and drives brown adipogenesis both *in vitro* and *in vivo*, indicating that *GLIS1* was a novel pro-adipogenic transcription factor (Tosic et al., 2018). Most recently, *GLIS1* was detected as a candidate gene of selective signature of sheep tail phenotype (Luo et al., 2021). A non-synonymous point mutation (g.27807636G > T) was found within *GLIS1* in two fat-tailed Chinese indigenous sheep breeds (Mongolian sheep and Small Tail Han sheep) compared with two thin-tailed dairy sheep (DairyMeade and East Friesian), and resulted in a Pro to Thr substitution (Luo et al., 2021). In our study, *GLIS1* was also strongly selected in Bamei Mutton sheep compared with Mongolian sheep. Taken together, *GLIS1*, as a pro-adipogenic factor, may play a key role in mesodermal cell differentiation during fetal development in fat-tailed sheep to initiate differentiation of pre-adipocytes and fat accumulation (Luo et al., 2021).

Nuclear receptor interacting protein one gene (*NRIP1*, also known as *RIP140*), encodes a nuclear protein also known as receptor-interacting protein 140 (*RIP140*). *RIP140* is widely expressed and plays an important role in regulating lipid and glucose metabolism (Leonardsson et al., 2004; Ho et al., 2011; Hochberg et al., 2015). *RIP140* interacts with multiple adipocyte-specific genes, such as *uncoupling protein 1* (*UCP1*), mitochondrial fatty acid transporter *carnitine palmitoyl transferase 1* (*CPT1*) and lipid droplet protein *cell death-inducing DFFA-like effector A* (*CIDEA*). The expression of these genes is characteristic of brown adipose tissue (Nautiyal et al., 2013). Previous studies in adipocyte cell models also revealed that *RIP140* functions as a corepressor of catabolic pathways, including fatty acid oxidation, oxidative phosphorylation, glycolysis and tricarboxylic acid cycle (Christian et al., 2005; Powelka et al., 2006). Moreover, Xu et al. conducted a genome-wide association study using phenotypes and genotypes of two breeds of contrasting tail types (Small-tailed and Large-tailed Han sheep breeds) to identify functional genes and variants associated with fat deposition, and revealed that *RIP140* was a strong candidate for fat deposition in the tails of sheep (Xu et al., 2017), which is consistent with our results.

It has been proved that androgen receptor (*Ar*) gene gets participate in lipid binding (Huang et al., 2009), and has a negative function in fat deposition in both mice and human beings (Rubinow et al., 2015; Kim et al., 2019). Adipose tissue macrophages express the androgen receptor (*AR*) and regulate adipose tissue remodeling. Thus, testosterone signaling in macrophages could alter the paracrine function of these cells and thereby contribute to the metabolic effects of androgens in men (Rubinow et al., 2015). In order to determine whether the loss of *AR* signaling in hematopoietic cells results in greater fat accumulation, Rubinow et al. performed a metabolic phenotyping study in male mice. C57BL/6J male mice (ages 12–14 weeks) underwent bone marrow transplant from either wild-type (WT) or *AR* knockout (*ARKO*) donors (n = 11–13 per group). Mice were fed a high-fat diet (60% fat) for 16 weeks. At baseline, 8 and 16 weeks, glucose and insulin tolerance tests were performed, and body composition was analyzed with fat-water imaging by MRI. No differences in body weight were observed between mice transplanted with WT bone marrow [WT (WTbm)] or *ARKO* bone marrow [WT (*ARKObm*)] prior to initiation of the high-fat diet. After 8 weeks of high-fat feeding, WT (*ARKObm*) mice exhibited significantly more visceral and total fat mass than WT (WTbm) animals. Resultant data indicate that *AR* signaling in hematopoietic cells influences body fat distribution in male mice, and the absence of hematopoietic *AR* plays a permissive role in visceral fat accumulation. These findings demonstrate a metabolic role for *AR* signaling in marrow-derived cells and suggest a novel mechanism by which androgen deficiency in men might promote increased adiposity (Rubinow et al., 2015). Kim et al. also discovered that blocking *AR* can decrease the expression of *CPT1* (one of long-chain fatty acid (LCFA) transport proteins) in the skeletal muscle, which reduces fat metabolism. Thus, reducing sex hormones or suppressing the sensitivity of *AR* can inhibit energy efficiency and fat metabolism by suppressing *CPT1* (Kim et al., 2019). However, the effect of *AR* gene on fat deposition in sheep is rarely reported. In our study, *AR* was identified as a promising candidate gene of sheep tail phenotype, and the same results were found in two Iranian thin- and fat-tailed

sheep Breeds (Moradi et al., 2022). Combined with the above studies, we speculate that *AR* may promote energy efficiency and fat metabolism of sheep, thereby inhibit fat deposition in tail.

Fibroblast growth factor 9 (*FGF9*) is a protein-coding gene that plays an important role in the regulation of embryonic development, cell proliferation, cell differentiation, and cell migration (Moioli et al., 2015). *FGF9* expressed not only in white adipose tissue (WAT) of human, but also in brown adipose tissue (BAT) while exposed to cold, regulating the development of adipose tissue (Mejert et al., 2010; Grefhorst et al., 2015). In mammals, WAT stores fat and BAT dissipates fat to produce heat. *FGF9* was selected as a candidate gene associated with tail phenotype in both our and a recent study (Moioli et al., 2015), indicating that *FGF9* may plays a role in fat deposition and metabolism of sheep, but the mechanism underlying it is unclear.

In addition to fat deposition, the number of caudal vertebrae also affects tail phenotype by affecting the length of sheep tail. It has been reported that *Vertnin* (*VRTN*) gene is a key candidate gene associated with the variation of vertebral number in sheep and pigs (Duan et al., 2018; Zhang et al., 2017). As a transcriptional inhibitor, *VRTN* independently regulates the expression of *BMP2* gene in the dorsoventral axis through combination with its regulatory sequence, thereby enabling normal development of embryos along the dorsoventral axis (Shao et al., 2017). In a study of genome-wide scan of selection signatures of sheep tail phenotype, Mastrangelo et al. identified *VRTN* as a key candidate gene; and the tail of the five fat-tailed sheep breeds where this signature was detected was definitively longer than the tail of the thin-tailed breeds (Mastrangelo et al., 2019). The same results were also found in Moioli et al., Zhu et al., and our study. In the current study, the tail length of Bamei Mutton sheep is longer than that of Mongolian sheep, with the caudal vertebrae number is 20–30 and 10–14, respectively. These results indicated that *VRTN* may be an important gene involved in regulating the development of sheep caudal vertebrae, and affect the tail phenotype of sheep by affecting the tail length.

Except for above mentioned genes, the rest of the candidate genes may also play important roles in regulating sheep tail phenotype, although there was not enough evidence at present based on publicly available information. Therefore, further studies and experiments are needed to confirm the roles and mechanisms of these genes in formation of sheep tail phenotype. In conclusion, our results provide a strong foundation for studying the regulation of tail phenotype in sheep and do offer hope that the causal mutations and the mode of inheritance of this trait will soon be discovered by further experimentation.

Data availability statement

The data presented in the study are deposited in the SRA repository, accession number PRJNA1192386 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1192386>).

Ethics statement

The animal studies were approved by the Animal Ethics Committee of Inner Mongolia Academy of Agricultural and

Animal Husbandry Sciences. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

YQ: Funding acquisition, Writing–original draft, Writing–review and editing, Methodology, Visualization. XH: Methodology, Resources, Writing–review and editing. BW: Data curation, Methodology, Writing–review and editing. CY: Visualization, Writing–review and editing, Data curation. LD: Resources, Writing–review and editing. BL: Data curation, Methodology, Validation, Visualization, Writing–review and editing. WZ: Conceptualization, Writing–review and editing. SF: Funding acquisition, Methodology, Writing–review and editing. YL: Conceptualization, Funding acquisition, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Modern Agricultural Industry Technology System of China (CARS-38), the Doctoral Start-up Fund Project of Xichang University (YBZ202406), the Inner Mongolia Autonomous Region Open Competition Projects (2022JBGS0024), and the Inner Mongolia Autonomous Region Science and Technology Plan (2023YFHH0114).

Conflict of interest

Author BL was employed by Inner Mongolia BIONEWS Technology Co., 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.2024.1509177/full#supplementary-material>

References

- Abdullah, A. Y., Kridli, R. T., Shaker, M. M., and Obeidat, M. D. (2010). Investigation of growth and carcass characteristics of pure and crossbred Awassi lambs. *Small Ruminant Res.* 94, 167–175. doi:10.1016/j.smallrumres.2010.08.005
- Abouheif, M. A., Kraides, M. S., and Shatat, R. A. (1993). Performance and carcass traits of docked and intact fat-tailed Najdi lambs. *Asian-Australasian J. Animal Sci.* 6 (1), 135–138. doi:10.5713/ajas.1993.135
- Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* 9 (1), 813. Published 2018 Mar 6. doi:10.1038/s41467-018-03206-y
- Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinforma.* 12, 246. Published 2011 Jun 18. doi:10.1186/1471-2105-12-246
- Atti, N., Bocquier, F., and Khaldi, G. (2004). Performance of the fat-tailed Barbarine sheep in its environment: adaptive capacity to alter-nation of underfeeding and re-feeding periods. A review. *Animal Res.* 53, 165–176. doi:10.1051/animres:2004012
- Atti, N., and Mahouachi, M. (2011). The effects of diet, slaughter weight and docking on growth, carcass composition and meat quality of fat-tailed Barbarine lambs. A review. *A Rev. Trop. Anim. Health Prod.* 43 (7), 1371–1378. doi:10.1007/s11250-011-9865-6
- Bicer, O., Pekel, E., and Guney, O. (1992). Effects of docking on growth performance and carcass characteristics of fat-tailed Awassi ram lambs. *Small Ruminant Res.* 8, 353–357. doi:10.1016/0921-4488(92)90216-q
- Bing, I. M., Aygn, T., kdal, G., and Yilmaz, A. (2006). The effects of docking on fattening performance and carcass characteristics in fat-tailed Norduz male lambs. *Small Ruminant Res.* 64 (1), 101–106. doi:10.1016/j.smallrumres.2005.04.011
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81 (5), 1084–1097. doi:10.1086/521987
- Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., et al. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science* 324 (5926), 532–536. doi:10.1126/science.1170587
- Christian, M., Kiskinis, E., Debevec, D., Leonardsson, G., White, R., and Parker, M. G. (2005). RIP140-targeted repression of gene expression in adipocytes. *Mol. Cell Biol.* 25 (21), 9383–9391. doi:10.1128/MCB.25.21.9383-9391.2005
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330
- Deng, J., Xie, X. L., Wang, D. F., Zhao, C., Lv, F. H., Li, X., et al. (2020). Paternal origins and migratory episodes of domestic sheep. *Curr. Biol.* 30 (20), 4085–4095.e6. doi:10.1016/j.cub.2020.07.077
- Dong, K., Yang, M., Han, J., Ma, Q., Han, J., Song, Z., et al. (2020). Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. *BMC Genomics* 21 (1), 800. doi:10.1186/s12864-020-07210-9
- Duan, Y., Zhang, H., Zhang, Z., Gao, J., Yang, J., Wu, Z., et al. (2018). VRTN is required for the development of thoracic vertebrae in mammals. *Int. J. Biol. Sci.* 14 (6), 667–681. Published 2018 Apr 30. doi:10.7150/ijbs.23815
- Eck, K., Kunz, E., Mendel, C., Luhken, G., and Medugorac, I. (2019). Morphometric measurements in lambs as a basis for future mapping studies. *Small Ruminant Res.* 181, 57–64. doi:10.1016/j.smallrumres.2019.04.007
- Grefhorst, A., van den Beukel, J. C., van Houten, E. L., Steenbergen, J., Visser, J. A., and Themmen, A. P. (2015). Estrogens increase expression of bone morphogenetic protein 8b in brown adipose tissue of mice. *Biol. Sex. Differ.* 6, 7. Published 2015 Apr 3. doi:10.1186/s13293-015-0025-y
- Ho, P. C., Chuang, Y. S., Hung, C. H., and Wei, L. N. (2011). Cytoplasmic receptor-interacting protein 140 (RIP140) interacts with perilipin to regulate lipolysis. *Cell Signal* 23 (8), 1396–1403. doi:10.1016/j.cellsig.2011.03.023
- Hochberg, I., Tran, Q. T., Barkan, A. L., Saltiel, A. R., Chandler, W. F., and Bridges, D. (2015). Gene expression signature in adipose tissue of acromegaly patients. *PLoS One* 10 (6), e0129359. Published 2015 Jun 18. doi:10.1371/journal.pone.0129359
- Holsinger, K. E., and Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat. Rev. Genet.* 10 (9), 639–650. doi:10.1038/nrg2611
- Huang, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211
- Kalds, P., Luo, Q., Sun, K., Zhou, S., Chen, Y., and Wang, X. (2021). Trends towards revealing the genetic architecture of sheep tail patterning: promising genes and investigatory pathways. *Anim. Genet.* 52 (6), 799–812. doi:10.1111/age.13133
- Kashan, N. E. J., Azar, G. H. M., Afzalzadeh, A., and Salehi, A. (2005). Growth performance and carcass quality of fattening lambs from fat-tailed and tailed sheep breeds. *Small Ruminant Res.* 60, 267–271. doi:10.1016/j.smallrumres.2005.01.001
- Khaldari, M., Kashan, N. E., Afzalzadeh, A., and Salehi, A. (2008). Growth and carcass characteristics of crossbred progeny from lean-tailed and fat-tailed sheep breeds. *South Afr. J. Animal Sci.* 37, 51–56. doi:10.4314/sajas.v37i1.4026
- Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., et al. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4 (3), e4668. doi:10.1371/journal.pone.0004668
- Kim, J., Park, J., Kim, N., Park, H. Y., and Lim, K. (2019). Inhibition of androgen receptor can decrease fat metabolism by decreasing carnitine palmitoyl transferase I levels in skeletal muscles of trained mice. *Nutr. Metab. (Lond)*. 16, 82. Published 2019 Nov 27. doi:10.1186/s12986-019-0406-z
- Kim, Y. S., Lewandoski, M., Perantoni, A. O., Kurebayashi, S., Nakanishi, G., and Jetten, A. M. (2002). Identification of Glis1, a novel Gli-related, Kruppel-like zinc finger protein containing transactivation and repressor functions. *J. Biol. Chem.* 277 (34), 30901–30913. doi:10.1074/jbc.M203563200
- Kridli, R., and Said, S. (1999). Libido testing and the effect of exposing sexually naïve Awassi rams to estrous ewes on sexual performance. *Small Rumin. Res.* 32 (2), 149–152. doi:10.1016/s0921-4488(98)00168-0
- Leonardsson, G., Steel, J. H., Christian, M., Pocock, V., Milligan, S., Bell, J., et al. (2004). Nuclear receptor corepressor RIP140 regulates fat accumulation. *Proc. Natl. Acad. Sci. U. S. A.* 101 (22), 8437–8442. doi:10.1073/pnas.0401013101
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27 (21), 2987–2993. doi:10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, Q., Lu, Z., Jin, M., Fei, X., Quan, K., Liu, Y., et al. (2020a). Verification and analysis of sheep tail type-associated PDGF-D gene polymorphisms. *Anim. (Basel)* 10 (1), 89. Published 2020 Jan 6. doi:10.3390/ani10010089
- Li, X., Yang, J., Shen, M., Xie, X. L., Liu, G. J., Xu, Y. X., et al. (2020b). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* 11 (1), 2815. Published 2020 Jun 4. doi:10.1038/s41467-020-16485-1
- Luo, R., Zhang, X., Wang, L., Zhang, L., Li, G., and Zheng, Z. (2021). GLIS1, a potential candidate gene affect fat deposition in sheep tail. *Mol. Biol. Rep.* 48 (5), 4925–4931. doi:10.1007/s11033-021-06468-w
- Marai, I. F. M., Nowar, M. S., Bahgat, L. B., and Owen, J. B. (1987). Effect of docking and shearing on growth and carcass traits of fat-tailed Ossimi sheep. *J. Agric. Sci.* 109 (3), 513–518. doi:10.1017/S0021859600081727
- Mastrangelo, S., Moiola, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F., et al. (2019). Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Prod. Sci.* 59, 835–848. doi:10.1071/an17753
- Mejert, N., Galitzky, J., Pettersson, A. T., Bambace, C., Blomqvist, L., Bouloumié, A., et al. (2010). Mapping of the fibroblast growth factors in human white adipose tissue. *J. Clin. Endocrinol. Metab.* 95 (5), 2451–2457. doi:10.1210/jc.2009-2049
- Moharrery, A. (2007). Effect of docking and energy of diet on carcass fat characteristics in fat-tailed Badghisian sheep. *Small Ruminant Res.* 69, 208–216. doi:10.1016/j.smallrumres.2005.12.021
- Moioli, B., Pilla, F., and Ciani, E. (2015). Signatures of selection identify loci associated with fat tail in sheep. *J. Anim. Sci.* 93 (10), 4660–4669. doi:10.2527/jas.2015-9389
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrabak, M., Dodds, K. G., Brauning, R., and McEwan, J. C. (2022). Hitchhiking mapping of candidate regions associated with fat deposition in Iranian thin and fat tail sheep breeds suggests new insights into molecular aspects of fat tail selection. *Anim. (Basel)* 12 (11), 1423. Published 2022 May 31. doi:10.3390/ani12111423
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrabak, M., Dodds, K. G., and McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* 13, 10. Published 2012 Feb 26. doi:10.1186/1471-2156-13-10
- Nakashima, M., Tanese, N., Ito, M., Auerbach, W., Bai, C., Furukawa, T., et al. (2002). A novel gene, GliH1, with homology to the Gli zinc finger domain not required for mouse development. *Mech. Dev.* 119 (1), 21–34. doi:10.1016/s0925-4773(02)00291-5
- Nautiyal, J., Christian, M., and Parker, M. G. (2013). Distinct functions for RIP140 in development, inflammation, and metabolism. *Trends Endocrinol. Metab.* 24 (9), 451–459. doi:10.1016/j.tem.2013.05.001
- Nekrutenko, A., and Taylor, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat. Rev. Genet.* 13 (9), 667–672. doi:10.1038/nrg3305
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8 (11), 857–868. doi:10.1038/nrg2187

- O'Donovan, P. B., Ghadaki, M. B., Behesti, R. D., Saleh, B. A., and Rollinson, D. H. L. (1973). Performance and carcass composition of docked and control fat-tailed Kellakui lambs. *Animal Sci.* 16, 67–76. doi:10.1017/s0003356100034887
- Olson, L. E., and Soriano, P. (2011). PDGFR β signaling regulates mural cell plasticity and inhibits fat development. *Dev. Cell* 20 (6), 815–826. doi:10.1016/j.devcel.2011.04.019
- Orihuela, A., and Ungerfeld, R. (2019). Tail docking in sheep (*Ovis aries*): a review on the arguments for and against the procedure, advantages/disadvantages, methods, and new evidence to revisit the topic. *Livest. Sci.* 230, 103837. doi:10.1016/j.livsci.2019.103837
- Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., et al. (2019). Rapid evolution of a retro-transposable hotspot of ovine genome underlies the alteration of BMP2 expression and development of fat tails. *BMC Genomics* 20 (1), 261. Published 2019 Apr 2. doi:10.1186/s12864-019-5620-6
- Pourlis, A. F. (2011). A review of morphological characteristics relating to the production and reproduction of fat-tailed sheep breeds. *Trop. Anim. Health Prod.* 43 (7), 1267–1287. doi:10.1007/s11250-011-9853-x
- Powell, A. M., Seth, A., Virbasius, J. V., Kiskinis, E., Nicoloso, S. M., Guilherme, A., et al. (2006). Suppression of oxidative metabolism and mitochondrial biogenesis by the transcriptional corepressor RIP140 in mouse adipocytes. *J. Clin. Invest* 116 (1), 125–136. doi:10.1172/JCI26040
- 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
- Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464 (7288), 587–591. doi:10.1038/nature08832
- Rubinow, K. B., Wang, S., den Hartigh, L. J., Subramanian, S., Morton, G. J., Buaas, F. W., et al. (2015). Hematopoietic androgen receptor deficiency promotes visceral fat deposition in male mice without impairing glucose homeostasis. *Andrology* 3 (4), 787–796. doi:10.1111/andr.12055
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007a). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449 (7164), 913–918. doi:10.1038/nature06250
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007b). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi:10.1038/nature06250
- Safarian, M., Zamiri, M. J., Hashemi, M., and Noorolahi, H. (2008). Relationships of fat-tail dimensions with fat-tail weight and carcass characteristics at different slaughter weights of Torkei-Ghashghai sheep. *Meat Sci.* 80 (3), 686–689. doi:10.1016/j.meatsci.2008.03.007
- Shao, M., Wang, M., Liu, Y. Y., Ge, Y. W., Zhang, Y. J., and Shi, D. L. (2017). Vegetally localized Vrtm functions as a novel repressor to modulate *bmp2b* transcription during dorsoventral patterning in zebrafish. *Development* 144 (18), 3361–3374. doi:10.1242/dev.152553
- Shelton, M., Willingham, T., Thompson, P., and Roberts, E. M. (1991). Influence of docking and castration on growth and carcass traits of fat-tail Karakul, Rambouillet and crossbred lambs. *Small Ruminant Res.* 4, 235–243. doi:10.1016/0921-4488(91)90147-i
- Sun, T., Huang, G. Y., Wang, Z. H., Teng, S. H., Cao, Y. H., Sun, J. L., et al. (2020). Selection signatures of Fuzhong Buffalo based on whole-genome sequences. *BMC Genomics* 21 (1), 674. Published 2020 Sep 29. doi:10.1186/s12864-020-07095-8
- Szpiech, Z. A., and Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31 (10), 2824–2827. doi:10.1093/molbev/msu211
- Tosic, M., Allen, A., Willmann, D., Lepper, C., Kim, J., Duteil, D., et al. (2018). Lsd1 regulates skeletal muscle regeneration and directs the fate of satellite cells. *Nat. Commun.* 9 (1), 366. Published 2018 Jan 25. doi:10.1038/s41467-017-02740-5
- Wang, F., Shao, J., He, S., Guo, Y., Pan, X., Wang, Y., et al. (2022). Allele-specific expression and splicing provide insight into the phenotypic differences between thin- and fat-tailed sheep breeds. *J. Genet. Genomics* 49 (6), 583–586. doi:10.1016/j.jgg.2021.12.008
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603
- Wang, Y. Q., Zhong, R. Z., Fang, Y., and Zhou, D. W. (2018). Influence of tail docking on carcass characteristics, meat quality and fatty acid composition of fat-tail lambs. *Small Ruminant Res.* 162, 17–21. doi:10.1016/j.smallrumres.2017.09.005
- Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., et al. (2015). Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16 (1), 194. Published 2015 Mar 17. doi:10.1186/s12864-015-1384-9
- Weir, B. S., and Hill, W. G. (2002). Estimating F-statistics. *Annu. Rev. Genet.* 36, 721–750. doi:10.1146/annurev.genet.36.050802.093940
- Xu, S. S., Ren, X., Yang, G. L., Xie, X. L., Zhao, Y. X., Zhang, M., et al. (2017). Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (*Ovis aries*). *Anim. Genet.* 48 (5), 560–569. doi:10.1111/age.12572
- 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 (1), 76–82. doi:10.1016/j.ajhg.2010.11.011
- Yousefi, A. R., Kohram, H., Zare Shahneh, A., Nik-khah, A., and Campbell, A. W. (2012). Comparison of the meat quality and fatty acid composition of traditional fat-tailed (Chall) and tailed (Zel) Iranian sheep breeds. *Meat Sci.* 92 (4), 417–422. doi:10.1016/j.meatsci.2012.05.004
- Yuan, Z., Liu, E., Liu, Z., Kijas, J. W., Zhu, C., Hu, S., et al. (2017). Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* 48 (1), 55–66. doi:10.1111/age.12477
- Zeng, K., Shi, S., and Wu, C. I. (2007). Compound tests for the detection of hitchhiking under positive selection. *Mol. Biol. Evol.* 24 (8), 1898–1908. doi:10.1093/molbev/msm119
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875
- Zhang, Z., Sun, Y., Du, W., He, S., Liu, M., and Tian, C. (2017). Effects of vertebral number variations on carcass traits and genotyping of Vertnin candidate gene in Kazakh sheep. *Asian-Australas J. Anim. Sci.* 30 (9), 1234–1238. doi:10.5713/ajas.16.0959
- 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. *Anim. (Basel)* 10 (5), 773. Published 2020 Apr 29. doi:10.3390/ani10050773
- Zhi, D., Da, L., Liu, M., Cheng, C., Zhang, Y., Wang, X., et al. (2018). Whole genome sequencing of hulunbuir short-tailed sheep for identifying candidate genes related to the short-tail phenotype. *G3 (Bethesda)* 8 (2), 377–383. doi:10.1534/g3.117.300307
- Zhu, C., Li, N., Cheng, H., and Ma, Y. (2021). Genome wide association study for the identification of genes associated with tail fat deposition in Chinese sheep breeds. *Biol. Open* 10 (5), bio054932. doi:10.1242/bio.054932



OPEN ACCESS

EDITED BY

Fei Hao,
Northumbria University, United Kingdom

REVIEWED BY

Quratulain Hanif,
University of Management and Technology,
Pakistan
Chaoyun Yang,
Xichang College, China

*CORRESPONDENCE

Xinhai Li
✉ zhumei185@163.com

RECEIVED 01 September 2024

ACCEPTED 14 November 2024

PUBLISHED 11 December 2024

CITATION

Yuan X, Meng K, Wang Y, Wang Y, Pan C,
Sun H, Wang J and Li X (2024) Unlocking the
genetic secrets of Dorper sheep: insights into
wool shedding and hair follicle development.
Front. Vet. Sci. 11:1489379.
doi: 10.3389/fvets.2024.1489379

COPYRIGHT

© 2024 Yuan, Meng, Wang, Wang, Pan, Sun,
Wang and Li. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Unlocking the genetic secrets of Dorper sheep: insights into wool shedding and hair follicle development

Xiaochun Yuan¹, Ke Meng¹, Yayan Wang¹, Yifan Wang¹,
Cuili Pan¹, Haoran Sun¹, Jankui Wang² and Xinhai Li^{1*}

¹College of Animal Science and Technology, Ningxia University, Yinchuan, China, ²Beijing Key Laboratory of Animal Genetic Improvement, China Agricultural University, Beijing, China

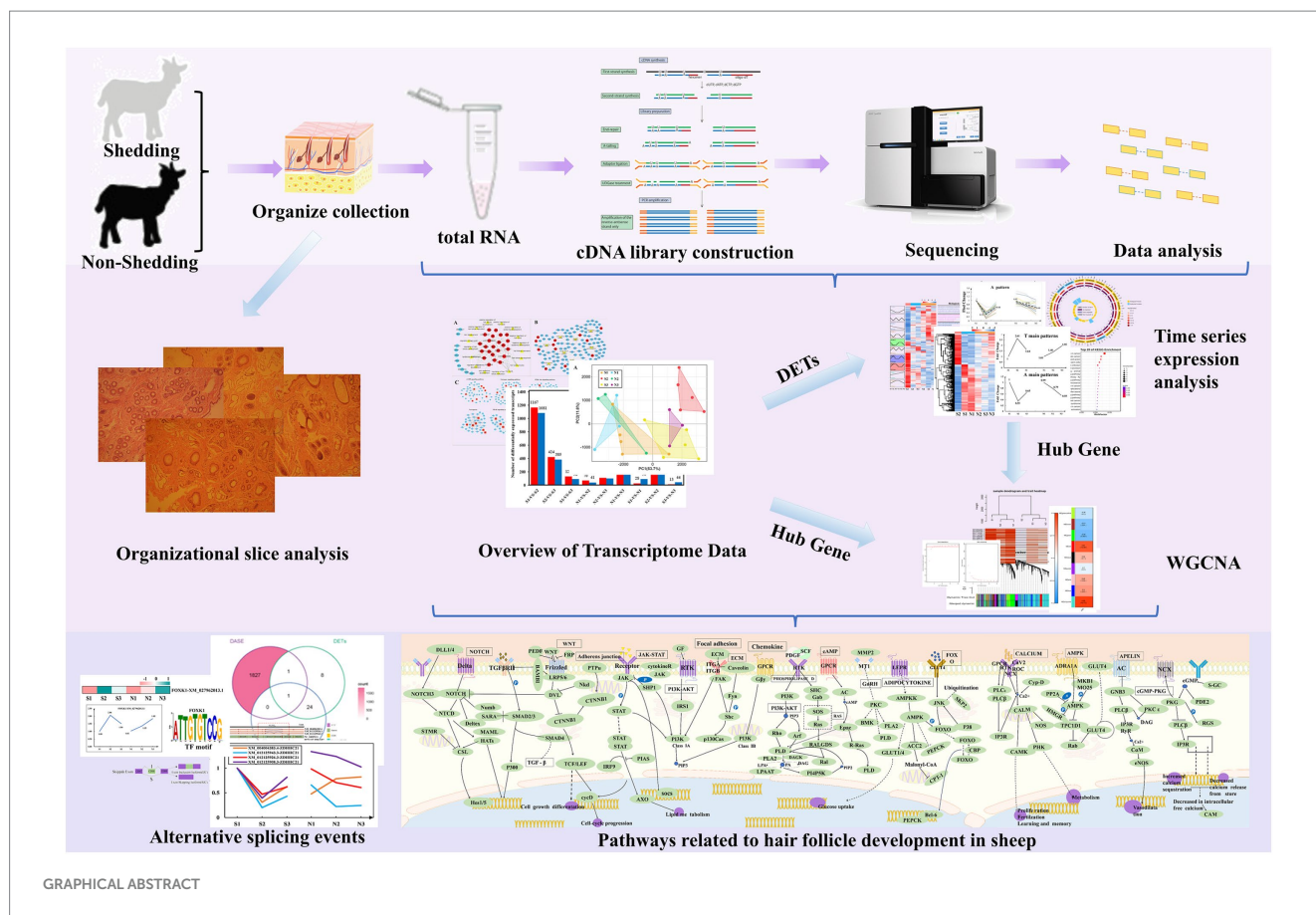
Dorper sheep is popular among farming enterprises with strong adaptability, disease resistance, and roughage tolerance, and an unique characteristic of natural shedding of wool. In a large number of observations on experimental sheep farms, it was found that the wool of some sheep still had not shed after May, thus manual shearing was required. Therefore, understanding the molecular mechanisms of normal hair follicles (HFs) development is crucial to revealing the improvement of sheep wool-related traits and mammalian skin-related traits. In this study, transcriptome analysis was performed on skin tissues of adult Dorper ewes in the shedding (S) and non-shedding (N) groups in September 2019, January 2020, and March 2020, respectively. The results identified 3,278 differentially expressed transcripts (DETs) in the three comparison groups within the S group, 720 DETs in the three comparison groups within the N group, and 1,342 DETs in the three comparison groups between the S-vs-N groups. Time-series expression analysis revealed 2 unique expression patterns in HF development, namely, elevated expression in the anagen phase (A pattern) and the telogen phase (T pattern). DETs with stage-specific expression had a significant presence in processes related to the hair cycle and skin development, and several classic signaling pathways involved in sheep HF development, such as Rap1, estrogen, PI3K-Akt, and MAPK, were detected. Combined analysis of DETs, time-series expression data, and weighted gene co-expression network analysis identified core genes and their transcripts influencing HF development, such as *DBI*, *FZD3*, *KRT17*, *ZDHHC21*, *TMEM79*, and *HOXC13*. Additionally, alternative splicing analysis predicted that the isoforms XM_004004383.4 and XM_012125926.3 of *ZDHHC21* might play a crucial role in sheep HF development. This study is a valuable resource for explaining the morphology of normal growth and development of sheep HFs and the genetic foundation of mammalian skin-related traits. It also offers potential insights into factors influencing human hair advancement.

KEYWORDS

Dorper sheep, RNA-Seq, hair follicles, transcript, WGCNA, alternative splicing events

Introduction

The hair follicles (HFs) of sheep constitute a “micro-organ” of the skin (1, 2). They are essential for sebum and sweat secretion, temperature regulation, homeostasis maintenance, and the regeneration and repair of skin stem cells (2). Additionally, HFs are an important source of tissues, which promote wool growth and determine the periodic growth characteristics of hair (3). Sheep HFs comprise primary hair follicles (PHFs) with medullated



hair and secondary hair follicles (SHFs) without medullated hair. Sheep wool is derived from the SHF structures in the skin (4, 5). The HF of sheep and wool growth and development follow a cyclic pattern of anagen, catagen, and telogen phases, which are influenced mainly by external environmental factors and specific genetic factors (6). The Wnt signaling pathway plays a crucial role in epidermal development, HF morpho-genesis, and regeneration (7). Gab1, which is downstream of receptor tyrosine kinase and upstream of Shp2 and Mapk, is involved in the regulation of the HF cycle and self-renewal of follicular stem cells (8). In normal skin, BMP controls epidermal homeostasis, HF growth, and melanogenesis (9). The APC (exon skipping), *POFUT1* (intron retention), and *TGFBR3* (cassette exon) genes, which are associated with HF development, are regulated by selective splicing (10). Transcription factors such as *KLF4*, *LEF1*, *HOXC13*, *RBPI*, *VDR*, *RARA*, and *STAT3* are specifically expressed in HF morphogenesis (11). Moreover, wool growth is a seasonal and cyclic phenomenon in animals that is controlled by photoperiod, inhibitory signaling, and the endocrine system (12, 13).

Interactions of multiple signaling pathways and specific genetic factors regulate HF growth and development cycle. The Wnt and Hippo signaling pathways promote Cashmere goat HF growth, whereas the Rap1, PI3K-Akt, NF-kappaB, and cAMP signaling pathways are crucial in the catagen and telogen phases. The PI3K-Akt signaling pathway and extracellular matrix (ECM) receptor interactions play important roles in transitioning from the telogen to the anagen phase, serving as candidate biomarkers for the regeneration process (14). *FGF5*, *FGFRL1*, and *RRAS* genes affect HF development through the MAPK signaling pathway in the Inner Mongolian velvet

goat (5). Vimentin regulates the growth cycle of Cashmere goat HF by affecting the outer root sheath (15). A comparison of molecular signals that initiate HF development on days 45, 55, and 65 in the embryos in Cashmere goats revealed significant differences in the Wnt, TGF- β , FGF, Hedgehog, and NOTCH signaling pathways from E55 to E65. *FOXN1*, *GATA3*, and *DLX3* may have consistent effects on HF development (10). The Notch signaling pathway plays a crucial role in HF differentiation and maturation, with regulatory factors such as *FOXN1*, *HOXC3*, *PRR13*, and *LHX15* potentially having consistent regulatory roles in Cashmere goat HF development (16). *KRT25*, *KRT27*, *KRT19*, *KRT10*, *KRT77*, *KRT1*, *KRT24*, *KRT14*, and *KRT4* are considered to be markers of the HF cycle (17). Additionally, the genes *LAMA5*, *WNT10A*, *KRT25*, *ZDHHC21*, *FZD1*, *LRP4*, *TGF β 2*, *TMEM79*, *SOX10*, *ITGB4*, and *KRT14* were significantly enriched in epidermal differentiation and development, hair follicle development, and hair follicle morphogenesis, and they were expressed specifically in the wool follicles of sheep at different developmental stages (18). In general, research on hair follicle development in cashmere goats has been widely reported, but research on wool follicles in sheep is relatively rare. In addition, the automatic shedding feature of Dorper sheep saves a certain amount of production costs for breeding companies, while research on some Dorper sheep that do not automatically shed their hair is even less common.

Weighted Correlation Network Analysis (WGCNA) allows for the joint analysis of high-throughput and phenotypic data, categorizing genes with similar expression profiles in modules and defining key modules and genes (19, 20). WGCNA has been used to identify key genes such as *WNT10A*, *KRT14*, *WNT11*, *LEF1*, *WNT5A*, *KRT1*, and

KRT6, which are associated with the development of the HF cycle in Inner Mongolian velvet goats. Among these, *WNT10A* is a crucial gene that regulates the development of HFs (18, 21, 22). ECM-receptor interactions, adhesion patches, the PI3K-Akt signaling pathway, and estrogen signaling pathways are closely related to the cyclic development of HFs. Genes such as *COL1A1*, *CIQTNF6*, and *COL1A2* play a role in increasing cashmere production (23). In addition, alternative splicing, as an important regulatory modality, has an intrinsic transcriptional regulatory mechanism that results in polymorphisms in transcript and protein structure and function. *FGF5* signaling is expressed at a much higher level in the final phase of the anagen phase than in the telogen phase, inducing a transition to the catagen phase. In contrast, *FGF5* signaling plays an inhibitory role in the anagen-to-catagen transition in the dermal cells of Cashmere goats (24, 25).

Numerous studies have demonstrated that wool follicle development is regulated by various pathways and specific genes throughout the growth cycle. However, most studies have focused on wool follicles and wool growth in Cashmere goats, with few studies addressing the entire cycle of wool follicle development and growth in sheep. Moreover, there are relatively few reports of Dorper sheep not shedding automatically. Additionally, regulation of the expression of single-gene multiple transcripts generated from mRNA precursors (pre-mRNAs) through different splicing modes during the cyclic changes in wool follicle development and growth in sheep have not been extensively studied, and the regulatory mechanisms that are involved remain largely unexplored. Furthermore, research on the impact of alternative splicing events on HF development in sheep is limited. In this study, transcript expression datasets were analyzed to identify key transcripts, pathways, and variable splicing events affecting the regulatory mechanisms of normal hair follicle development and growth cycles in sheep. Our study aims to provide a theoretical foundation to elucidate the regulatory mechanisms underlying the cyclic wool follicle growth in sheep.

Materials and methods

Animals and sample collection

All adult Dorper ewes used in this study were obtained from Zhong mu Yilin Livestock Co. Ltd. (Yinchuan, China). Sustained observations at the experimental sheep farm revealed that Dorper sheep exhibited a trait-separation phenomenon between shedding and non-shedding wool. By May, some sheep continued to retain their wool and required manual shearing. Consequently, Dorper adult ewes with consistent feeding and management conditions, good body condition, and similar age (2 years) were selected for this study. Ewes were categorized into a shedding group (S) and a non-shedding group (N) to investigate the cyclic development of HFs. Initially, 10 Dorper sheep were selected from the S group and the N group as the experimental model. Five Dorper sheep from the group S and three from the group N with the best phenotype were finally selected for subsequent analysis. The samples were collected in September 2019, January 2020, and March 2020. Skin tissues were collected from the same batch of sheep in group S (5 sheep) and group N (3 sheep) at each time point. Samples were collected from the posterior edge of the last rib at the junction of the rib and the midline of the body (Sample

size: 2cm²). For sampling, the wool is first sheared as far as possible. Next, the sampling site was disinfected and injected with an anesthetic drug. Then, after the skin tissue was cut off using scissors, the sheep wounds were sprayed with iodophor. Finally, anti-inflammatory drugs were applied to the wounds, and the wounds were sutured. We have done our best to minimize the suffering of the sheep. Tissue samples were rinsed with phosphate-buffered saline, immediately transferred to RNAase-free cryo-preservation tubes, and preserved in liquid nitrogen. The stages of Group S were labeled as S1, S2, and S3, whereas those of Group N were labeled as N1, N2, and N3 according to the sampling time. Additionally, skin tissues were collected for paraffin sectioning.

Histological staining

After immersing sheep skin samples in 4% paraformaldehyde for 24 h for fixation, the samples were embedded and rinsed for 30 min with running water to eliminate the fixative and then dehydrated in gradient ethanol solutions. The tissue was embedded in paraffin using a JB-P5 machine. Paraffin blocks were sectioned along the horizontal axis using an RM2016 microtome (Germany) to yield Sections approximately 5 µm thick. The sections were photographed at 40× magnification using an OLYMPUS cellSens (BX53) microscope. PHF and SHF diameters were measured using ImageJ software.¹ Data analysis was conducted using one-way ANOVA with the Agricol package (v.1.3–7) for R studio software.

Total RNA extraction and transcriptome sequencing

Total RNA was isolated from 24 skin samples using TRIzol reagent (Invitrogen, United States). RNA purity, concentration, and integrity of each sample were checked with a Nanodrop 2000 and an Agilent 2,100 Bioanalyzer with RIN ≥ 8.4, and 1.8 < optical density_{260/280} < 2.0. An RNA-seq library was constructed using a TruSeq™ RNA sample preparation kit (Illumina; San Diego, CA) following the manufacturer's instructions. First, ribosomal RNA was removed to maximize the retention of all coding RNAs. The resulting RNA was then randomly fragmented into short pieces of approximately 300 bp. Subsequently, these RNA fragments served as templates for synthesizing the first strand of cDNA using six-base random hexamers. Next, the second strand of cDNA was synthesized by adding buffer, dNTPs (with dUTP instead of dTTP), RNase H, and DNA polymerase I, followed by purification using a QiaQuick polymerase chain reaction (PCR) kit and elution with EB buffer for terminal repair, the addition of base A, and sequencing adapter. Degradation of the second chain was achieved using the enzyme uracil-N-glycosylase. Lastly, PCR amplification and fragment size selection were performed using agarose gel electrophoresis. The constructed sequencing library was then sequenced on the Illumina HiSeq™ 4,000 platform with 150 bp paired-end reads. Each sample generates approximately 6 Gb of sequencing data.

¹ <https://imagej.net/ij/download.html>

Data quality control, comparison, assembly, and expression calculation

The following analysis was done based on MobaXterm (v.23.1). Quality filtering of raw data was performed using FastQC (v.0.11.9) software to remove reads containing adapter sequences, reads with >10% N content, duplicate reads, and low-quality reads (where bases with a quality value $Q \leq 20$ made up more than 50% of the total reads), to obtain clean reads. HISAT2 (v.2.2.1)² (26) was used to align the reads to the sheep reference genome (Oar_rambouillet_v1.0), and StringTie (v.2.2.0) (27) software was used to assemble the transcripts and calculate the transcript expression in each sample. The expression level is displayed as raw reads count and FPKM. The raw reads count indicates the number of reads contained in the transcript, but due to the influence of sequencing amount and transcript length, the raw reads count is not conducive to the comparison of differential transcripts between samples. To ensure the accuracy of subsequent analysis, we first corrected the sequencing depth, then corrected the length of the transcript, and obtained the FPKM value of the transcript before conducting subsequent analysis.

Bioinformatics analysis

Sample relationship analysis

The following analysis was performed based on R studio. In order to understand the repeatability between samples and assist in excluding abnormal samples. The gmodels package (v.2.19.1) was used to perform principal component analysis (PCA). The corrplot package (v.0.92) was used to perform a Pearson correlation analysis on all samples.

Identification of differential transcripts

Normalization, p -value calculation, and multiple hypothesis testing corrections (false discovery rate, FDR) were conducted using the DESeq2 (v.1.42.0) package on each sample's read count (28). Differentially expressed transcripts (DETs) were identified using $|\log_2FC| \geq 1$ and $P\text{-adjust} < 0.05$ as the criteria. Quantitative relationships of DETs among groups were plotted using GraphPad (v.8.0.1), and the intersecting genes among the DETs were visualized using the UpSetR package (v.1.4.0).

Enrichment function analysis and network visualization

Functional annotation and enrichment of all transcript from sequencing were performed using Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) databases. Gene network interactions were visualized using Cytoscape (v3.9.1) (29).

Time series expression analysis

The heatmap package (v.1.0.12) was used to visualize the clustering among samples, and the STEM software³ was used to

classify the expression patterns of DETs and identify the expression patterns and key transcripts related to HF development.

Weighted correlation network analysis

Co-expression networks were constructed using the WGCNA (v.1.72–5) package. Use the pickSoftThreshold function to select the appropriate soft threshold power β value to calculate the network structure. Merging modules with 75% feature transcript similarity using a dynamic tree-cutting technique to ascertain the final module count. Transcript with $|MM| \geq 0.75$ in key modules were correlated with hub transcript (Demonstrated as a gene in the main text) related to HF development in sheep to pinpoint the hub transcript (Demonstrated as a gene in the main text) and their regulatory relationships with HF development pathways. In addition, sequence motif enrichment in the promoters of transcripts specifically expressed during the stages related to HF development was analyzed using MEME (v. 5.5.5).

Alternative splicing analysis

Alternative splicing is a crucial gene regulatory mechanism in eukaryotes. RMATS (30) was used to determine 5 variable splicing events, namely, exon skip (ES), retained intron (RI), mutually exclusive exons (MXEs), alternative 5' splice site (A5SS), and alternative 3' splice site (A3SS) (31). Differential alternative splicing events were screened using the criteria of $|\text{IncLevelDiff}| \geq 0.1$, $p\text{-value} \leq 0.05$, and $\text{FDR} \leq 0.05$. Key candidate genes related to HF development in sheep were identified via combined analysis of differential alternative splicing events, DETs, and hub genes. The expression, splicing types, and locations of these candidate genes were analyzed in depth. Motifs in the splice site sequences of skipped exons with significant splicing changes were identified using the software XSTREME (v.5.5.5) (32). These splice site sequences replaced the 500 bp downstream sequence of the exon (exonStart-exonEnd: 89851506–89,851,692).

RT-qPCR

Six DETs (*MT1C*, *KRT27*, *TPRX1*, *FABP9*, *KRT23*, and *S100A14*) were randomly selected, and RT-qPCR was used to determine their relative expression. Exon sequences of the relevant transcripts were downloaded from The National Center for Biotechnology Information (NCBI) database, and primers were designed using Primer Premier 6.0, with primer specificity verified using NCBI Primer-BLAST. GAPDH served as the internal reference gene, with primers synthesized by Bioengineering Co. (Shanghai; Supplementary Table S1). cDNA was synthesized from the extracted total RNA using a PrimeScript™ RT kit (Takara, China). qRT-PCR was performed using a TB Green® Premix Ex Taq™ Fluorescence Quantification kit to determine transcript expression. The qRT-PCR reaction system included the following: 10 μL TB Green Premix Ex Taq II (Tli RNaseH Plus), 0.8 μL forward primer (10 $\mu\text{mol/L}$), 0.8 μL reverse primer (10 $\mu\text{mol/L}$), 0.4 μL ROX reference dye (50 $\times$), 2 μL cDNA, and 6 μL RNase-free water, made up to a total volume of 20 μL . The real-time fluorescence quantitative PCR procedure was as follows: pre-denaturation at 95°C for 30 s; denaturation at 95°C for 10 s, annealing at 57.8°C for 30 s, and extension at 65°C for 5 s, for a total of 40 cycles. Data were organized using Microsoft Excel (v.2019), and

² <https://daehwankimlab.github.io/hisat2/>

³ www.omicshare.com/tool

relative transcript expression was calculated using the $2^{-\Delta\Delta C_t}$ method. GraphPad software was used to analyze the graphs.

Results

Histological changes in the HFs of sheep

Skin samples were sectioned along the horizontal axis to observe the histological changes in sheep wool follicles during September 2019, January 2020, and March 2020 (Figures 1A,B). In group S, PHFs and SHFs transitioned from the anagen phase to the telogen phase between S1 and S2, with SHFs appearing smaller or even atrophied. Both PHFs and SHFs from S2 to S3 showed signs of developing hair buds and re-entered growth and development. In group N, the PHFs in group N2 still had developing hair buds, and the SHFs did not

exhibit significant shrinkage compared with those in group S2. PHFs and SHFs in groups N2 to N3 were in a slow anagen phase. These results indicated that HFs in group S exhibited a rhythmic cyclic development, whereas those in group N lacked a telogen phase, resulting in a continuous growth and development process without a cyclic pattern. Next, we standardized the diameter measurements (μm) of PHFs and SHFs in groups S and N. Highly significant differences were noted between the PHFs in both groups S and N at all 3 time points. However, SHFs were significantly different only in January and not significantly different in either September or March (Figure 1C; Supplementary Table S2). Wool shedding refers to the cycle of growth and development of SHFs, shedding, and regrowth. Therefore, S1, N1, and N2 were initially defined as the anagen phase, S2 as the telogen phase, S3 as the early anagen phase, and N3 as the slow anagen phase. Additionally, the observation of tissue sections provided a reliable basis for selecting samples from specific HF

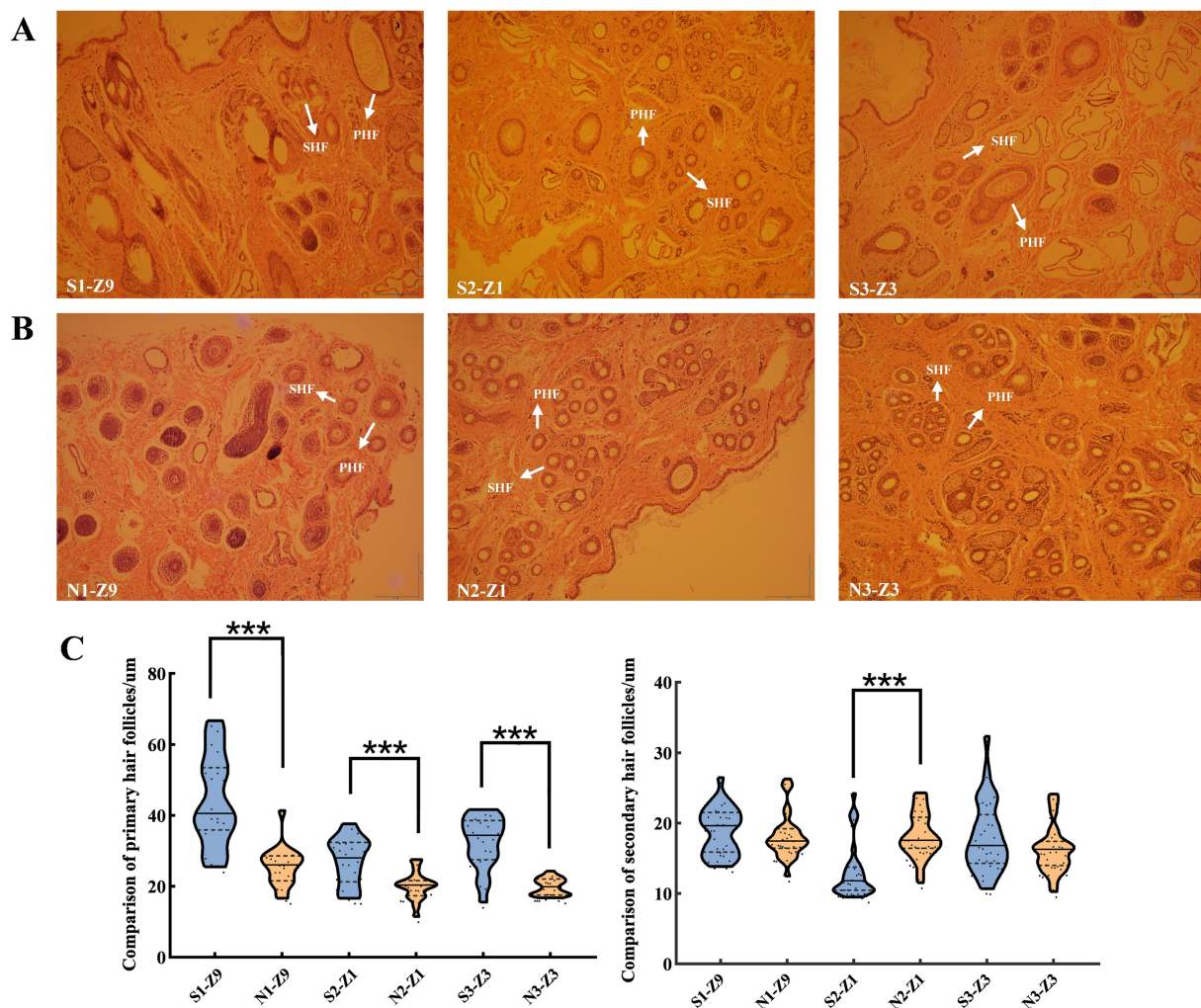


FIGURE 1

Histological comparison of sheep HFs at 3 developmental stages. (A) and (B) display the histological changes in HFs observed in September, January, and March, respectively (magnification: $\times 40$). (A) September (S1), January (S2), and March (S3). (B) September (N1), January (N2), and March (N3). (C) Measurement of HF diameter (μm) of PHFs and SHFs at 3 developmental stages of sheep HFs and one-way analysis of variance (mean $\pm$ standard deviation). 20 PHFs and 30 SHFs were selected for each sample to statistically analyze the mean and standard deviation. In total, the diameters of 120 PHFs and 180 SHFs were measured. Note: HF: hair follicle, PHF: primary hair follicle, and SHF: secondary hair follicle. *** $p < 0.001$ indicates extremely significant difference. Z1, Z3, and Z9 represent the sampling years. Z1, Z3, and Z9 represent the sampling years. Z9: September 2019, Z1: January 2020, Z3: March 2020.

developmental states for transcriptome sequencing and subsequent analysis to identify the transcript related to HF development in sheep.

Quality assessment results and sample correlation analysis

To investigate the regulatory mechanisms governing HF growth and development in sheep, a total of 24 samples from both groups S (5) and N (3) were analyzed. The experiment yielded 2,178,092,054 clean reads (Supplementary Table S3) and detected 47,660 transcripts through RNA-seq data analysis (Supplementary Table S4). PCA revealed that S1 and S3 samples were closer to each other, whereas S2 samples were farther from both S1 and S3. By combining the sampling time and section results, it was determined that S1 and S3 were in the anagen phase of HF development, whereas S2 was in the telogen phase. N1 and N2 clustered with S1 and N3 showed expression similar to S3, further suggesting that there was no telogen phase in the HFs of group N sheep. Therefore, it was determined that S1, N1, and N2 were in the anagen phase, S2 was in the telogen phase, S3 was in the early anagen phase, and N3 was in the slow anagen phase (Figure 2A). Correlation analysis was conducted on all samples, and the correlation of duplicate samples within the same treatment group was found to be higher than 0.95 (Figure 2B).

RT-qPCR validation of RNA-seq

Six DETs were randomly selected and their comparative expression was determined using RT-qPCR. Results from RT-qPCR for these 6 DETs were consistent with the transcript expression profiles, validating the reliability of the transcript sequencing results (Figure 3).

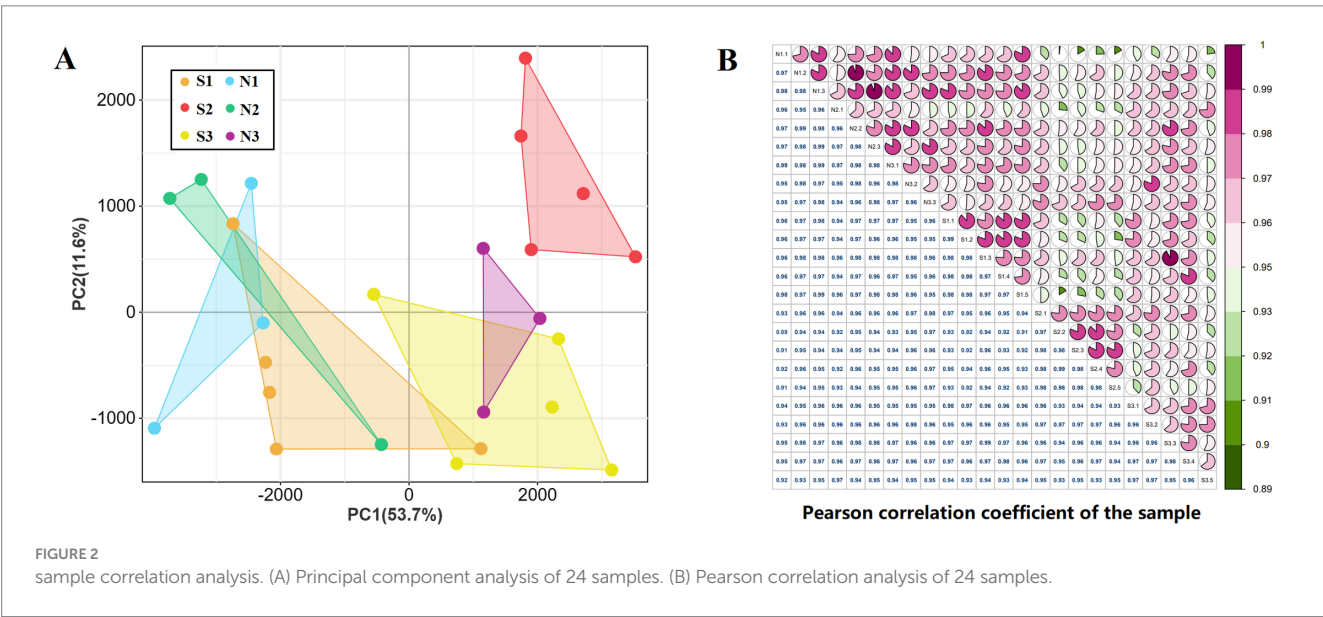
Identification of differential transcripts and their enrichment analysis.

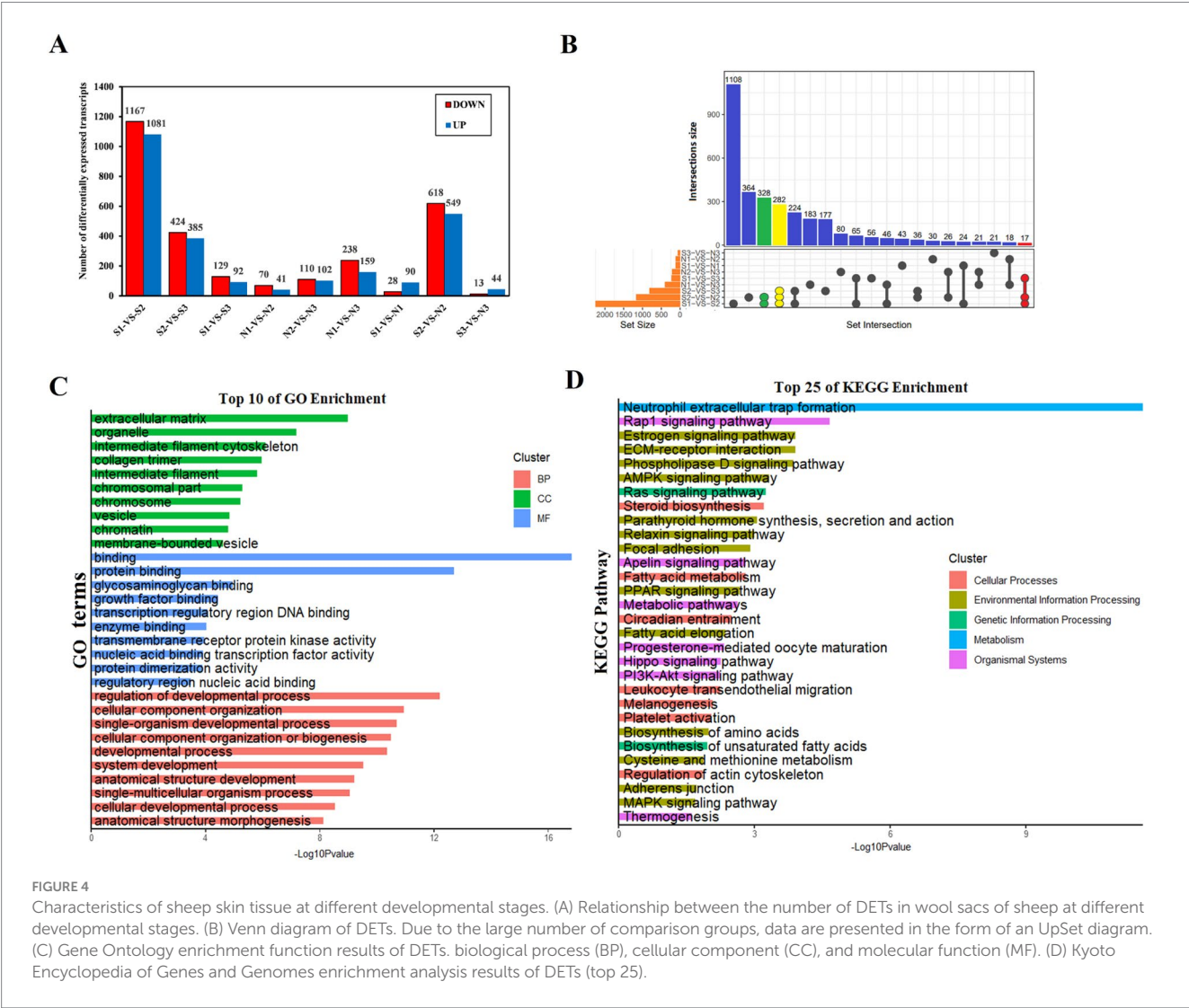
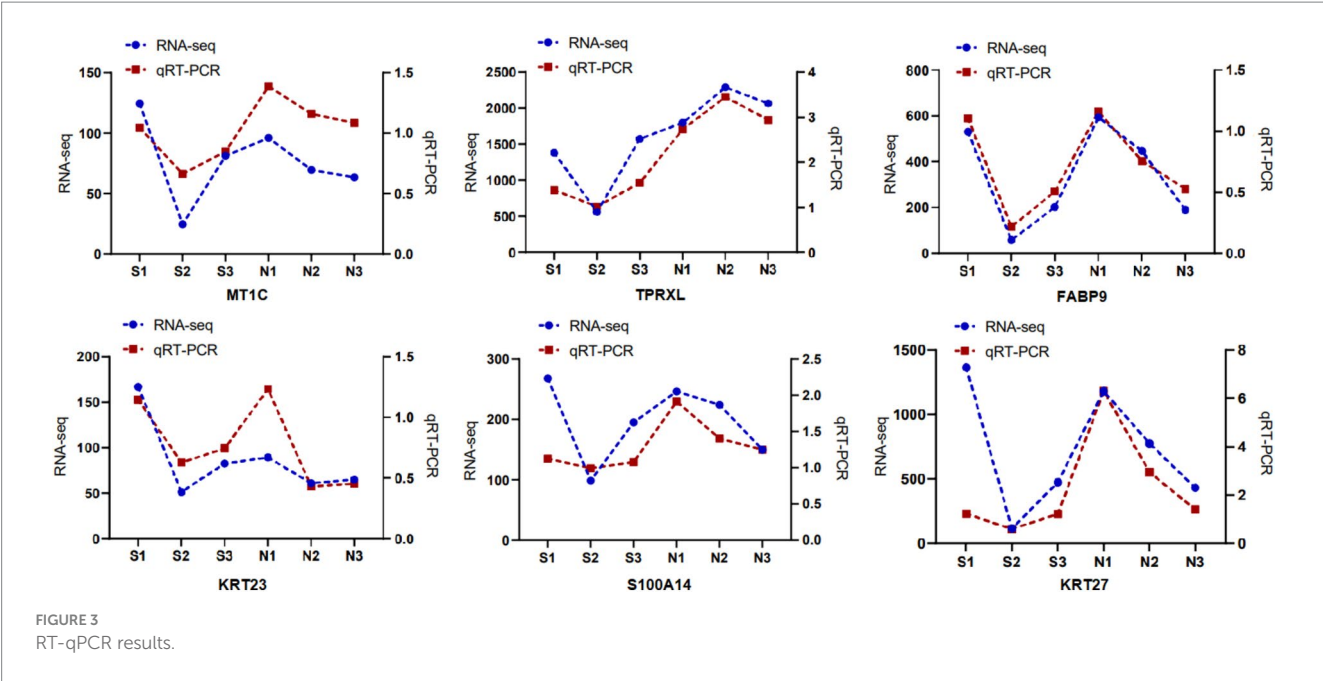
A total of 5,340 DETs were identified using the DESeq2 package with $|\log_2FC| \geq 1$ and P-adjust <0.05 as the screening criteria

(Figure 4A; Supplementary Table S5). These included 3,278 DETs for intra-Group S comparisons, 720 DETs for intra-Group N comparisons, and 1,342 DETs for comparisons between Groups S and N. Among the 9 groups that were compared, 3 DET sets were of particular interest, all intersecting with N2 and comprising 627 DETs (Figure 4B). GO and KEGG analyses were performed on 5,340 DETs. In the biological process (BP) category, DETs were mainly enriched in the regulation of developmental processes, cellular component organization, and single-organism developmental processes (Figure 4C; Supplementary Table S6). Within the CC category, DETs were predominantly associated with the extracellular matrix, organelles, and the intermediate filament cytoskeleton (Figure 4C; Supplementary Table S6). Within the MF category, DETs were chiefly related to binding, protein binding, and glycosaminoglycan binding (Figure 4C; Supplementary Table S6). DETs were significantly enriched in the pathways of alcoholism, systemic lupus erythematosus, neutrophil extracellular trap formation, Rap1, estrogen, ECM-receptor interaction, phospholipase D, AMPK, and Ras signaling (Figure 4D; Supplementary Table S7).

Enrichment analysis of hair follicle development-related transcripts

To further identify the biological functions associated with HF development in sheep and their potential key transcripts. Cytoscape software was used to visualize GO terms and KEGG pathways in HF development. Enriched BP terms included hair cycle, hair cycle process, regulation of the hair cycle, follicle development, negative regulation of the hair cycle, positive regulation of follicle maturation, positive regulation of HF development, regulation of HF maturation, HF maturation, dermal development, regulation of epidermal development, positive regulation of epidermal development, and negative regulation of epidermal development. Relevant transcripts include the 23 genes, namely, *KRT17*, *ROCK2*, *PRKCH*, *Ldb1*, *RBPI*, *DBI*, *ERRF1*, *DKK1*, *LRP4*, *ALOX15B*, *LAMA5*, *Trpv3*, *FZD3*, *MSX2*, *SOX9*, *HOXC13*, *NGFR*, *SMAD4*, *KRT84*, *TMEM79*, *LOC101108627*,





LOC101116039, and *LOC654331* (Figure 5A; Supplementary Table S8). Similarly, in pathway analyses, 13 genes (*KRT17*, *PRKCH*, *RBPJ*, *DBI*, *DKK1*, *FZD3*, *SMAD4*, *ALOX15B*, *LAMA5*, *NGFR*, *SOX9*, *ROCK2*, and *LOC101108627*) were found to be enriched in Wnt, Notch, PPAR, estrogen, metabolic pathway, cAMP, PI3K-Akt, necroptosis, and vascular smooth muscle contraction signaling pathways. Interestingly, alongside the genes associated with GO terms and KEGG pathways, various genes, including *LEF1*, *CTNNB1*, *WNT6*, *KRT25*, *WNT4*, *KRT40*, *KRT27*, and *BAMBI*, were associated with our traits of interest (Figure 5B). Analogous to the previous GO and KEGG analyses, GO and KEGG analyses were conducted on the pathways potentially related to sheep wool follicle growth and progression at various stages. Additionally, a gene network diagram for these pathways was constructed according to their respective stages (Figure 5C; Supplementary Figure S1). Our results matched those described above and also uncovered several genes in sheep that may be involved in HF

growth and development that interact with the above signaling pathways.

Dynamic expression patterns of DETs during HF development in sheep

A hierarchical cluster analysis of all DETs was performed to determine the expression patterns of DETs associated with HF development in sheep. This analysis revealed 2 major expression modules, namely, high and low expression during the S2 period (Figure 6A). Next, a time-series expression analysis was conducted on all DETs, which were classified into 10 expression modules based on changes in their expression patterns (Figure 6A). GO functional enrichment analysis was performed separately on these 10 modules, which identified 11 BP terms related to HF development in sheep

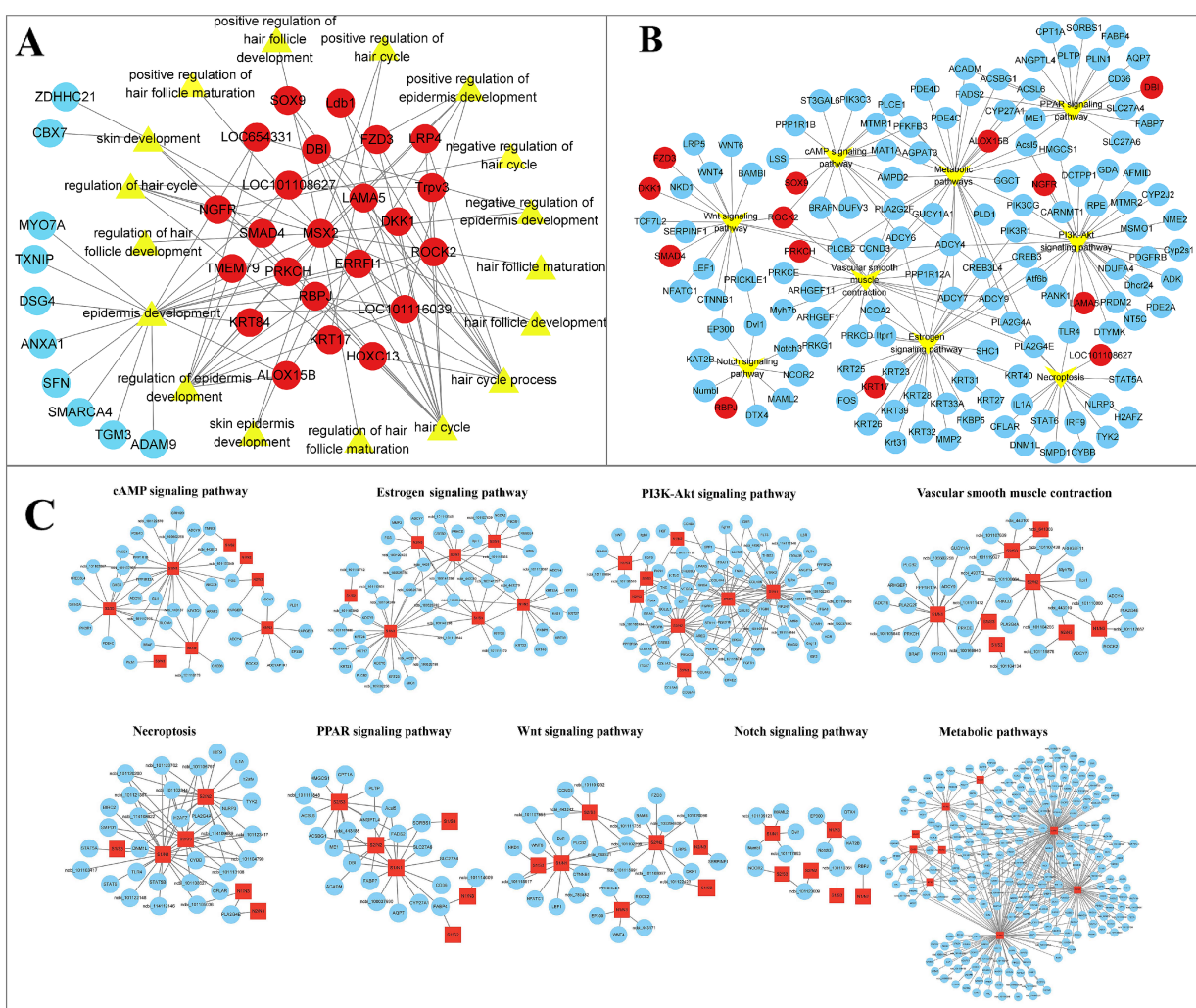


FIGURE 5
Visualization of GO terms and DETs connected to the development of HFs in sheep. (A) Gene network diagram of biological process (BP) terms during HF morphogenesis in sheep. Red indicates joint effects across multiple BP terms, whereas blue indicates effects in a single BP term. The transcript numbers corresponding to the genes are listed in Supplementary Table S8. (B) Network visualization of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and their associated genes during HF morphogenesis in sheep. (C) Occurrence of signaling pathways, including Wnt, Notch, PPAR, estrogen, metabolic pathway, cAMP, PI3K-Akt, necroptosis, and vascular smooth muscle contraction, during sheep wool follicle morphogenesis. GO and KEGG functional enrichment analyses on pathways potentially related to sheep HF growth and development at various phases. The visual network diagram is displayed in Supplementary Figure S1.



(Continued)

FIGURE 6 (Continued)

with the enriched terms related to the corresponding biological processes (Top 4) shown on the right. (B) Genes related to HF development in sheep are primarily concentrated in modules 2, 4, and 7 ($p < 0.05$). The corresponding biological processes, genes, and TFs are shown on the right. The transcript numbers corresponding to the genes are listed in [Supplementary Table 10](#). (C) Patterns A (intense expression throughout the anagen phase) and T (intense expression during the telogen phase). Bold dashed lines indicate the main modes (average) of A and T patterns. (D) Summary of the A and T pattern of all DETs. (E–G) TFs are mainly enriched in Gene Ontology terms (Top 20). Visualization using enrichment circles and z-score bubble charts. (F–H) TFs are mainly enriched in Kyoto Encyclopedia of Genes and Genomes terms (Top 20). Bubble charts and secondary classification bar charts are shown.

([Figure 6A](#); [Supplementary Table S9](#)). In Module 2, transcripts associated with HF development were enriched for processes including hair cycle, skin development, positive regulation of epidermis development, regulation of epidermis development, and epidermis development. These transcripts were characterized by low expression at S2 and high expression at N2 ([Figure 6B](#); [Supplementary Table S10](#)). Module 4 was associated with similar processes, including regulation of hair cycle, regulation of epidermis development, epidermis development, and positive regulation of epidermis development, with the transcripts also showing low expression at S2 ([Figure 6B](#); [Supplementary Table S10](#)). Analysis of the transcripts related to HF development in Modules 2 and 4 revealed that most showed strong expression in the anagen phase (S1), low expression during the telogen phase (S2), and a slight increase in expression during the early anagen phase (S3). In contrast, the N group exhibited a consistent decrease in expression (N1 and N2 in the anagen phase, and N3 in the slow anagen phase), which was defined as the A pattern ([Figure 6C](#)). Module 7 plays a key role in the hair cycle and hair cycle processes ([Figure 6B](#); [Supplementary Table S10](#)). Its expression pattern is the opposite of that observed in Modules 2 and 4 (low expression at S1, high expression at S2, and slightly reduced expression at S3). In the N group, the expression of N1, N2, and N3 increased consistently, and this was defined as the T pattern ([Figure 6C](#)).

Motif enrichment analysis was performed on the promoters (2000 bp upstream and 500 bp downstream of the transcription start site) of HF development-related transcripts in Modules 2, 4, and 7 to determine whether the stage-specifically expressed transcripts were co-regulated by certain TFs. Motifs for TFs were predominantly enriched in the Kruppel-like factor family, Sp family, and ZNF family ([Supplementary Table S11](#)). Analysis of the expression patterns of these TFs revealed significant differences between S2 and N2, suggesting that these TFs may play an important role in regulating normal HF growth cycles ([Figure 6B](#)). TFs showed significant enrichment in BPs, including the regulation of gene expression, cellular macromolecule biosynthetic process, macromolecule biosynthetic process, and gene expression ([Figures 6E–G](#)). They are also involved in the dynamic regulation of signaling pathways including TGF- β , GnRH, Estrogen, and Apelin pathways ([Figures 6F–H](#)). Additionally, for screening A and T pattern transcripts among all DETs, a total of 1,005 A and T pattern transcripts were identified, comprising 605 A pattern and 400 T pattern transcripts ([Figure 6D](#); [Supplementary Table S12](#)). KEGG analysis revealed that A pattern transcripts were primarily enriched in pathways related to systemic lupus erythematosus, alcoholism, neutrophil extracellular trap formation, estrogen signaling, necroptosis, and metabolic pathways ([Figure 6D](#); [Supplementary Figures S2A,B](#)), whereas T pattern transcripts were mainly enriched in pathways associated with salivary secretion, oxytocin signaling, Apelin signaling, circadian entrainment, cGMP-PKG signaling, and Rap1 signaling ([Figure 6D](#);

[Supplementary Figures S2C,D](#)). The clustering of A and T pattern transcripts within these signaling pathways suggests a close relationship between A and T pattern transcripts and HF development.

WGCNA analysis identifies stage-specific co-expression modules

The KRT Family Members are one of the main components of wool tissue structure and serve as an important marker for HF growth and development (17). The expression of 26 KRTs in the sequencing results was analyzed in this study. The overall trend of KRTs is shown as A pattern. The homogenization of the expression of 21 KRTs matching this pattern was calculated ([Figure 7A](#); [Supplementary Table S13](#)). Correlation analysis between KRT family as an indicator of hair follicle development status (Phenotypic data) and transcript expression. The results showed that groups S1, N1, and N2 had the highest correlation with the development status data, followed by groups S3 and N3, whereas group S2 showed the least correlation ([Figures 7B,C](#)). In this clustering, N2 was grouped with S1 and N1, indicating its closer association with the HF anagen period. The expression of KRTs in the S2 stage was significantly lower than that in the S3 stage (the early anagen phase), being only 24% of the anagen phase level, suggesting that S2 represents the telogen phase of HFs ([Figure 7A](#)). In summary, the developmental status of HFs in groups S and N was further confirmed, wherein S1, N1 and N2 corresponded to the anagen phase, S2 corresponded to the telogen phase, S3 corresponded to the early anagen phase, and N3 corresponded to the slow anagen phase. These results correspond with those from PCA and time-series expression analysis.

To construct the co-expression network of 24 samples, we chose $\beta = 5$ to construct the TOM clustering tree and merged similar modules (modules with 75% feature gene similarity), resulting in 9 modules ([Supplementary Figure S3](#)). The expression pattern heatmap showed that the MEturquoise ($\text{cor} = 0.95$, $p = 4e-12$), MEred ($\text{cor} = 0.81$, $p = 1e-06$), and MEgreen ($\text{cor} = -0.56$, $p = 0.004$) modules were significantly correlated with KRTs ([Figure 7D](#)). Therefore, transcripts from these 3 modules were chosen for GO (BP) and KEGG analyses. The red module was significantly enriched for negative regulation of HF maturation, regulation of the hair cycle, negative regulation of the hair cycle, and epidermis development. The green module was enriched for epithelial cell differentiation, epithelial cell–cell adhesion, and skin epidermis development related to HF development. The turquoise module was mainly enriched for positive regulation of epidermis development, positive regulation of HF development, and regulation of epidermis development ([Figure 7D](#); [Supplementary Table S14](#)). In combination with DETs and time-series expression analysis, the gene relationships between Wnt, Ras, PI3K-Akt, MAPK, focal adhesion, and Rap1 pathways using KEGG are shown in [Figure 7E](#) ([Supplementary Table S14](#)). Grounded in the candidate genes

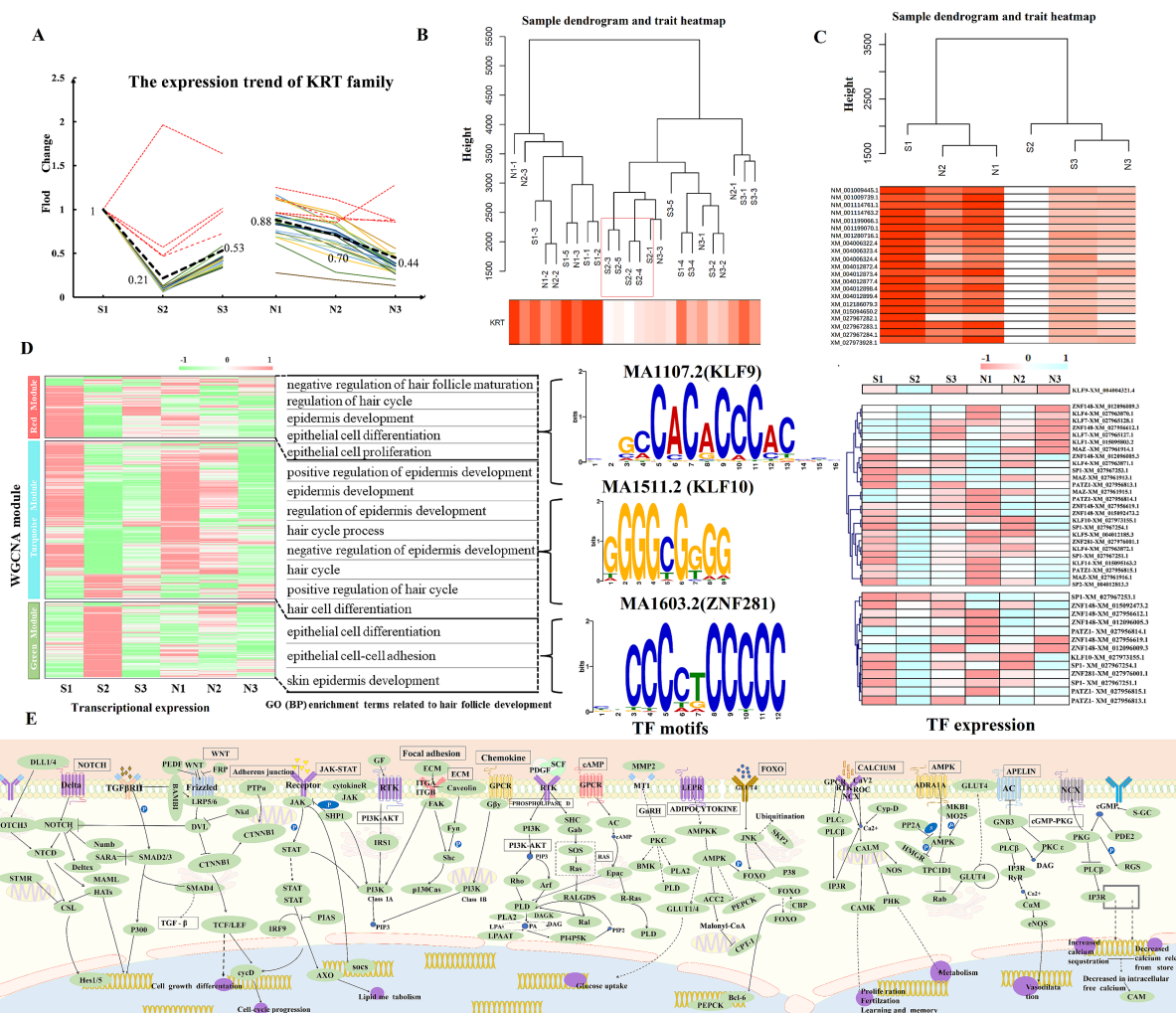


FIGURE 7

Weighted gene co-expression network analysis. Note: (A) Normalized expression pattern of the KRT family. The color indicates the expression pattern of 21 KRTs (A pattern). The bold, black dashed line indicates the main pattern of KRTs, whereas the red dashed line represents the deleted value, which does not appear in the subsequent analysis. (B,C) Hierarchical clustering tree of KRTs and sheep skin transcript expression. Figure B shows the clustering of the main pattern of KRTs and 24 samples, and Figure C shows the clustering of 21 KRTs and transcripts at each stage. (D) Expression clustering of transcripts in 3 key modules. Biological process terms related to HF development corresponding to the key modules are enriched on the right, with related TF genes and leading sequence motifs displayed beside each module. (E) Pathways related to HF development in sheep.

connected to HF development in sheep (Supplementary Table S15), 11 core genes ($|MM| \geq 0.75$), namely, *DBI*, *FZD3*, *KRT17*, *ZDHHC21*, *TGM3*, *DSG4*, *TMEM79*, *KRT84*, *HOXC13*, *LOC101116039*, and *MSX2* were identified in the red, green, and turquoise modules as being involved in the control of HF growth and development in sheep.

Analysis of alternative splicing events related to HF development in sheep

Alternative splicing is an important regulatory modality affecting HF development in sheep. To investigate its effect on the cyclic development of sheep HF, rMTAS software was used to analyze transcriptome data of sheep skin. Differential alternative splicing events were identified in the three transition phases (anagen-telogen, telogen-early anagen, early anagen-anagen) within the S group and the N group. All 6 comparison groups showed the most ES-type differential alternative splicing events (Figure 8A;

Supplementary Table S16). Among these, the alternative splicing events that were significantly different only in the transition periods of group S were of particular interest as they may be closely related to HF development in sheep (Figure 8B). GO (BP) enrichment function analysis of the 6 comparison groups showed that they were mainly significantly enriched in metabolic process regulation, negative regulation of actin filament depolymerization, and organization of ribonucleoprotein complex subunits. In addition, they were also significantly enriched in skin development and HF development (Figure 8C). Intersecting these results with the genes related to HF development identified the *ZDHHC21* (Figure 8D). Among its transcripts, XM_012125908.3, XM_004004383.4, XM_012125943.3, and XM_012125926.3 were consistently expressed in group S but abnormally expressed in group N (Figure 8D).

The prediction of transcription factors (TFs) for the *ZDHHC21* revealed that *FOXX1* (XM_027962013.1) exhibited an expression pattern opposite to that of the *ZDHHC21* transcript (Figure 8D). The alternative splicing events in *ZDHHC21* were further analyzed.

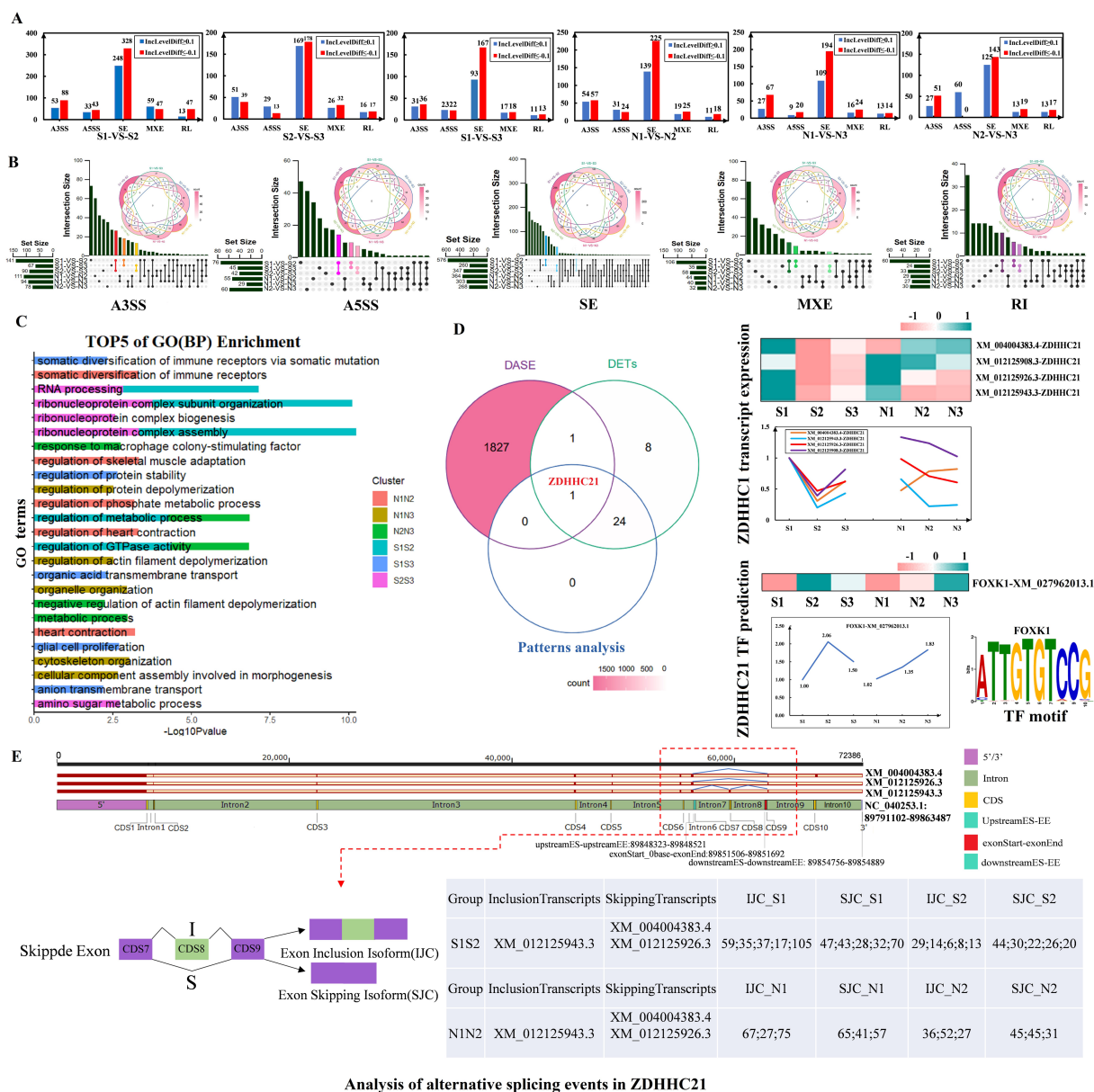


FIGURE 8
Analysis of alternative splicing events related to hair follicle development in sheep. (A) Analysis of differential alternative splicing events across 3 transition periods. (B) Venn diagram showing the 6 comparison groups. Alternative splicing events that are significantly different only in the S group transition period are highlighted. (C) Top 5 biological processes of the 6 comparison groups. (D) Venn diagram of differential alternative splicing events, differentially expressed transcripts and Pattern Analysis. The right side shows the expression of the *ZDHHC21* transcripts, related TFs, and the top representative sequence motifs. (E) Detailed analysis of alternative splicing positions of the *ZDHHC21* gene. IJC indicates the number of relevant reads that support the retention of the skipped exon (exon inclusion), which will appear in the mature mRNA after splicing. SJC records the number of relevant reads supporting the splicing of the skipped exon (exon skipping), which is excised during alternative splicing and does not appear in mature mRNA.

XM_004004383.4 and XM_012125926.3 of *ZDHHC21* exhibited exon skipping events at CDS8 (89854756–89,854,889), whereas XM_012125943.3 had an intron retention event at CDS8 (Figure 8E). In the S1-vs-S2 comparison, the SJC_S2 of *ZDHHC21* was nearly double that of IJC_S2. In contrast, in the N1-vs-N2 comparison, the difference between SJC_N2 and IJC_N2 was smaller, highlighting a significant contrast with the S1-vs-S2 comparison (Figure 8E). Combining the expression of the 3 transcripts, it was predicted that XM_004004383.4 and XM_012125926.3 of *ZDHHC21* might play a decisive role in HF development in sheep (Figures 8D,E). In summary,

there were notable differences in the reads of the final processed mRNA transcripts of *ZDHHC21* after alternative splicing in groups S and N, which may directly affect the growth and developmental status of HFs.

Discussion

In mammals, each mature HF functions as a regenerative system, enabling animal hair to grow in a regular pattern (33). Understanding

the morphological and molecular mechanisms underlying the normal growth and development of wool follicles in sheep is crucial for advancing our knowledge of hair growth biology and the genetic underpinnings of wool traits. In this analysis, we selected skin tissues from Dorper sheep in September, January, and March for histomorphological analysis between the shedding and non-shedding groups. In group S, HFs developed in a rhythmic cyclical manner, from anagen (S1) to telogen (S2) and re-entered the early anagen phase (S3). In group N, HFs in the telogen phase (N2) still exhibited growing and developing hair buds, indicating that the follicles did not follow a cyclic pattern and remained in a constant state of growth and development. Wool shedding was found to be a cyclic process involving the growth and development of SHFs, shedding of wool, and subsequent hair regrowth (4, 5, 34). Standardized diameter measurements (μm) of SHFs in groups S and N revealed highly significant differences in SHFs only in January (telogen phase), with no significant differences observed in September or March. This finding indicated that HFs in the N2 period were in a constant state of growth and development. Additionally, the growth and developmental state of HFs in group S was determined to be in the anagen phase in September, the telogen phase in January, and the early anagen phase in March, which was consistent with that in previous studies (12, 35–37).

From a transcriptome perspective, 3,278, 720, and 1,342 DETs were identified in the S, N, and S-vs-N groups. DETs were significantly enriched in signaling pathways such as the Rap1, estrogen, ECM-receptor interaction, phospholipase D, AMPK, NOTCH, MAPK, and Ras pathways. Previous studies have shown that the PI3K-Akt, MAPK, NOD-like receptor, ECM-receptor interaction and Rap1 pathways are involved in the growth and development of hair follicles in cashmere goats (14, 18). Additionally, the PPAR pathway plays a role in HF development, control of keratinocyte differentiation, and the development of functional skin barrier (38, 39). The MAPK signaling pathway is crucial in mammalian HFs and plays a role in keratinocyte differentiation, epidermal cell differentiation, and multicellular biological development (8, 40). High-throughput transcriptome sequencing identified differences in gene expression between primary and secondary hair follicles and showed that angiogenesis, ECM receptor interactions, and the Wnt/ β -catenin/Lef1 signaling pathway are closely associated with hair follicle morphogenesis (34). The Notch pathway plays a key role in determining cell fate, regulating the proliferation and differentiation of epidermal tissue cells, and wound healing (41, 42). Notch signaling plays a role in the late stage of embryonic hair follicle formation, and hair lacking Notch signaling will appear abnormal (43). In addition, during the development of hair follicles in mice, knocking out β -catenin prevents the formation of dermal condensates, underscoring the critical role of the LEF/TCF/ β -catenin signaling pathway in regulating dermal condensate formation (44–47). DKK1 inhibits the Wnt signaling pathway by blocking the phosphorylation of β -catenin, thereby inducing hair follicle regression (48). The PI3K/Akt signaling pathway is essential for maintaining and restoring the hair-inductive capacity of human dermal cells and promoting hair follicle regeneration (49, 50). Furthermore, the PI3K/Akt, Wnt, Notch, and BMP signaling pathways participate in the growth and development of hair follicles and skin tissues in Rex rabbits. These pathways also facilitate the transition of hair follicles from the anagen phase to the catagen and telogen phases, affecting hair density in Rex rabbits (51).

In this study, the Rap1, Ras, MAPK, phospholipase D, and estrogen signaling pathways were significantly enriched in the S1-vs-S2 comparison group, suggesting that these pathways and their associated genes may regulate the transition of HFs from the anagen to the telogen phase. In the S2-vs-S3 comparison group, the PPAR, ECM-receptor interaction, PI3K-Akt, AMPK, and focal adhesion signaling pathways were significantly enriched, indicating that genes in these pathways may play a role in the transition from the telogen to the early anagen phase, stimulating the generation of new hair shafts and leading to the timely shedding of old wool. The Rap1, estrogen, and Ras pathways were significantly re-enriched in the S3-vs-S1 comparison group, suggesting their crucial role in facilitating the cycling process of HFs from the early anagen phase to the anagen phase. The N1-vs-N2 comparison group did not show significant enrichment in pathways such as Rap1, estrogen, MAPK, and phospholipase D pathways compared with that in the S group. However, these pathways appeared in the N2-vs-N3 comparison group. The Rap1, estrogen, ECM-receptor interaction, and Ras pathways were significantly enriched in the N3-vs-N1 comparison group and were similar to the enrichment results noted in the S3-vs-S1 comparison group. These results indicated that the HFs in group N did not enter the telogen phase. Thus, it could be concluded that the HFs in group N did not follow a cyclic pattern and were always in a state of growth and development. The results of this study align with findings from previous research on cashmere goats, mice, and even human dermal cells. The PI3K-Akt signaling pathway and ECM-receptor interactions play important roles in the transition from the telogen to the anagen phase and serve as candidate biomarkers for this regeneration (14). Compared to our study, the PI3K/Akt pathway in Rex rabbits primarily facilitates the transition of hair follicles from the anagen phase to the telogen, contrary to our findings, suggesting that PI3K/Akt expression may vary across different mammals. Furthermore, The PI3K-Akt signaling pathway is crucial for regulating keratinocyte survival and proliferation, among other functions (52).

Using time series expression analysis and WGCNA, *MSX2* (53), *HOXC13* (54), *DBI*, *FZD3* (55), *ZDHHC21* (56), *PRKCH*, *LOC101116039*, *KRT17* (57), *DSG4* (58), *TMEM79* (59), and *LOC101108627* were determined to be associated with sheep hair follicle development. Additionally, *ALOX15B*, *LOC654331*, *SFN*, *Ldb1* (60), *DKK1* (61), *NGFR* (62), *LAMA5* (63), *SMAD4* (64), and *LRP4* (65, 66) were found to be associated with hair cycle processes, skin development, epidermis development, and hair cycle regulation. The above genes related to hair follicle development showed stage-specific expression, with *MSX2*, *HOXC13*, *KRT17*, *DBI*, *FZD3*, and *ZDHHC21* showing the highest expression during the anagen phase (A pattern). In contrast, *Ldb1*, *DKK1*, *NGFR*, *LAMA5*, *SMAD4*, and *LRP4* were highest expressed during the telogen phase (T pattern). In addition, The KRT family is a major component of wool, and the keratins *KRT25*, *KRT27*, *KRT19*, *KRT10*, *KRT77*, *KRT1*, and others are regarded as markers of the hair follicle cycle. They are significantly enriched in estrogen and ECM-receptor interaction signaling pathways and play a very important role in the development of secondary hair follicles (17). KRTs have been associated with the development of the hair follicle in yaks, and gene expression correlation analysis of the keratin family revealed a strong positive correlation of KRTs mainly throughout the hair follicle development cycle (67). TFs such as *TCF3*, *TCF4*, *ZNF740*, *EGR1*, *FLI1*, *SP1*, *E2F6*, and *ZNF148*, among others, were predicted to regulate these genes

associated with HF growth and development. Signaling pathways were significantly enriched with these TFs, as revealed by functional enrichment analysis such as the TGF- β , GnRH, estrogen, and apelin pathways. Previous studies have demonstrated that transcription factors such as *EGR1*, *LEF1*, *HOXC13*, *RBPJ*, *VDR*, *RARA*, and *STAT3* have stage-specific roles in HF morphogenesis (11). The study shows that *EGR1* is closely associated with HF growth and development and plays a crucial role in embryonic organ formation (68–71). The TFs *Lef1* and *Twist* are expressed in the early stages of dermal condensates (44, 47). The role of *CUX1* in *in vitro* dermal papilla cell proliferation may be regulated by *SP1* and *KROX20* (72). Additionally, *ZDHHC21* has been identified as the key gene affecting HF morphogenesis in Merino sheep (18). Members of the ZDHHC family mediate post-translational modifications via palmitoylation, with the *ZDHHC21* protein being specifically expressed in the skin and restricted to a particular hair lineage. Loss of *ZDHHC21* function results in delayed hair shaft differentiation at the gene expression level (56). In the current study, analysis of alternative splicing events in the *ZDHHC21* gene revealed a significant exon skipping event in the S1-vs-S2 comparison, with SJC_S2 being almost twice that of IJC_S2, unlike that in the N1-vs-N2 comparison. Based on the transcript expression of the *ZDHHC21* gene, we could predict that the isoforms XM_004004383.4 and XM_012125926.3 may play a crucial role in HF development in sheep. Additionally, alternative splicing, a key mode of transcriptional regulation, showed differences in the mRNA reads of the *ZDHHC21* isoforms between groups S and N, which may directly influence HF growth and development.

Conclusion

RNA-seq was used to sequence skin tissues from sheep HFs during the anagen, telogen, and early anagen phases, which led to the identification of key candidate genes and their corresponding transcripts associated with HF development in sheep. For example, *HOXC13*, *DBI*, *FZD3*, *ZDHHC21*, *KRT8*, and *SMAD4*. Several signaling pathways related to HF development, including the Rap1, Ras, MAPK, Jak-STAT, and PI3K-Akt pathways, were identified across the 3 HF development stages. Furthermore, it was found that the XM_004004383.4 and XM_012125926.3 transcripts of the *ZDHHC21* gene may play a decisive role in the development of HF in sheep. This study serves as a valuable resource for interpreting the morphology of normal HF growth and development in sheep, elucidating the genetic basis of mammalian skin-associated traits, and offering potential insights into the molecular mechanisms of human hair growth and development.

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 at: <https://www.ncbi.nlm.nih.gov/bioproject/>; PRJNA963059.

References

- Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. *Curr Biol*. (2009) 19:R132–42. doi: 10.1016/j.cub.2008.12.005
- Houschyar KS, Borrelli MR, Tapking C, Popp D, Puladi B, Ooms M, et al. Molecular mechanisms of hair growth and regeneration: current understanding and novel

Ethics statement

The animal study was approved by All sample collections followed the guidelines of the Experimental Animal Welfare Ethics Committee of Ningxia University (Animal Ethics Approval No. NXU-19-018). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

XY: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. KM: Investigation, Writing – review & editing. YaW: Investigation, Writing – review & editing. YiW: Data curation, Formal analysis, Writing – review & editing. CP: Writing – review & editing. HS: Writing – review & editing. JW: Conceptualization, Methodology, Supervision, Writing – review & editing. XL: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by the National Natural Science Foundation of China Screening of key molecules and construction of regulatory networks for sheep shedding traits (No. 31960650).

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.2024.1489379/full#supplementary-material>

- paradigms. *Dermatology (Basel, Switzerland)*. (2020) 236:271–80. doi: 10.1159/000506155
3. Galbraith H. Fundamental hair follicle biology and fine fibre production in animals. *Animal: Int J Animal Biosci*. (2010) 4:1490–509. doi: 10.1017/s17517311000025x
 4. Baker RE, Murray PJ. Understanding hair follicle cycling: a systems approach. *Curr Opin Genet Dev*. (2012) 22:607–12. doi: 10.1016/j.gde.2012.11.007
 5. Su R, Gong G, Zhang L, Yan X, Wang F, Zhang L, et al. Screening the key genes of hair follicle growth cycle in inner Mongolian cashmere goat based on RNA sequencing. *Archives animal breeding*. (2020) 63:155–64. doi: 10.5194/aab-63-155-2020
 6. Li C, He X, Wu Y, Li J, Zhang R, An X, et al. Single-cell transcriptome sequence profiling on the morphogenesis of secondary hair follicles in Ordos fine-wool sheep. *Int J Mol Sci*. (2024) 25:25. doi: 10.3390/ijms25010584
 7. Veltri A, Lang C, Lien WH. Concise review: Wnt signaling pathways in skin development and epidermal stem cells. *Stem cells (Dayton, Ohio)*. (2018) 36:22–35. doi: 10.1002/stem.2723
 8. Akilli Öztürk Ö, Pakula H, Chmielowiec J, Qi J, Stein S, Lan L, et al. Gab1 and Mapk signaling are essential in the hair cycle and hair follicle stem cell quiescence. *Cell Rep*. (2015) 13:561–72. doi: 10.1016/j.celrep.2015.09.015
 9. Botchkarev VA. Bone morphogenetic proteins and their antagonists in skin and hair follicle biology. *J Invest Dermatol*. (2003) 120:36–47. doi: 10.1046/j.1523-1747.2003.12002.x
 10. Zhang Y, Wang L, Li Z, Chen D, Han W, Wu Z, et al. Transcriptome profiling reveals transcriptional and alternative splicing regulation in the early embryonic development of hair follicles in the cashmere goat. *Sci Rep*. (2019) 9:17735. doi: 10.1038/s41598-019-54315-7
 11. Zhao B, Luo H, He J, Huang X, Chen S, Fu X, et al. Comprehensive transcriptome and methylome analysis delineates the biological basis of hair follicle development and wool-related traits in merino sheep. *BMC Biol*. (2021) 19:197. doi: 10.1186/s12915-021-01127-9
 12. Wang S, Ge W, Luo Z, Guo Y, Jiao B, Qu L, et al. Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (*Capra hircus*). *BMC Genomics*. (2017) 18:767. doi: 10.1186/s12864-017-4145-0
 13. Foldes A, Hoskinson RM, Baker P, McDonald BJ, Maxwell CA, Restall BJ. Effect of immunization against melatonin on seasonal fleece growth in feral goats. *J Pineal Res*. (1992) 13:85–94. doi: 10.1111/j.1600-079X.1992.tb00059.x
 14. Li C, Feng C, Ma G, Fu S, Chen M, Zhang W, et al. Time-course RNA-seq analysis reveals stage-specific and melatonin-triggered gene expression patterns during the hair follicle growth cycle in *Capra hircus*. *BMC Genomics*. (2022) 23:140. doi: 10.1186/s12864-022-08331-z
 15. Rile N, Liu Z, Gao L, Qi J, Zhao M, Xie Y, et al. Expression of vimentin in hair follicle growth cycle of inner Mongolian cashmere goats. *BMC Genomics*. (2018) 19:38. doi: 10.1186/s12864-017-4418-7
 16. Gao Y, Wang X, Yan H, Zeng J, Ma S, Niu Y, et al. Comparative transcriptome analysis of fetal skin reveals key genes related to hair follicle morphogenesis in cashmere goats. *PLoS One*. (2016) 11:e0151118. doi: 10.1371/journal.pone.0151118
 17. Gao WZ, Xue HL, Yang JJSRR. Proteomics analysis of the secondary hair follicle cycle in Liaoning cashmere goat. *Small Rumin Res*. (2021) 201:106408. doi: 10.1016/j.smallrumres.2021.106408
 18. He J, Zhao B, Huang X, Fu X, Liu G, Tian Y, et al. Gene network analysis reveals candidate genes related with the hair follicle development in sheep. *BMC Genomics*. (2022) 23:428. doi: 10.1186/s12864-022-08552-2
 19. Galán-Vázquez E, Perez-Rueda E. Identification of modules with similar gene regulation and metabolic functions based on co-expression data. *Front Mol Biosci*. (2019) 6:139. doi: 10.3389/fmolb.2019.00139
 20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinfo*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
 21. Wu Z, Hai E, Di Z, Ma R, Shang F, Wang Y, et al. Using WGCNA (weighted gene co-expression network analysis) to identify the hub genes of skin hair follicle development in fetus stage of Inner Mongolia cashmere goat. *PLoS One*. (2020) 15:e0243507. doi: 10.1371/journal.pone.0243507
 22. Gong G, Fan Y, Yan X, Li W, Yan X, Liu H, et al. Identification of genes related to hair follicle cycle development in Inner Mongolia cashmere goat by WGCNA. *Front Vet Sci*. (2022) 9:894380. doi: 10.3389/fvets.2022.894380
 23. Wang J, Sui J, Mao C, Li X, Chen X, Liang C, et al. Identification of key pathways and genes related to the development of hair follicle cycle in cashmere goats. *Gene*. (2021) 12:12. doi: 10.3390/genes12020180
 24. He X, Chao Y, Zhou G, Chen Y. Fibroblast growth factor 5-short (FGF5s) inhibits the activity of FGF5 in primary and secondary hair follicle dermal papilla cells of cashmere goats. *Gene*. (2016) 575:393–8. doi: 10.1016/j.gene.2015.09.034
 25. Imamura T. Physiological functions and underlying mechanisms of fibroblast growth factor (FGF) family members: recent findings and implications for their pharmacological application. *Biol Pharm Bull*. (2014) 37:1081–9. doi: 10.1248/bpb.b14-00265
 26. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. (2015) 12:357–60. doi: 10.1038/nmeth.3317
 27. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. (2015) 33:290–5. doi: 10.1038/nbt.3122
 28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8
 29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303
 30. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA*. (2014) 111:E5593–601. doi: 10.1073/pnas.1419161111
 31. Park E, Pan Z, Zhang Z, Lin L, Xing Y. The expanding landscape of alternative splicing variation in human populations. *Am J Hum Genet*. (2018) 102:11–26. doi: 10.1016/j.ajhg.2017.11.002
 32. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res*. (2015) 43:W39–49. doi: 10.1093/nar/gkv416
 33. Balañá ME, Charreau HE, Leirós GJ. Epidermal stem cells and skin tissue engineering in hair follicle regeneration. *World J Stem Cells*. (2015) 7:711–27. doi: 10.4252/wjcs.v7.i4.711
 34. Zhu B, Xu T, Yuan J, Guo X, Liu D. Transcriptome sequencing reveals differences between primary and secondary hair follicle-derived dermal papilla cells of the cashmere goat (*Capra hircus*). *PLoS One*. (2013) 8:e76282. doi: 10.1371/journal.pone.0076282
 35. Yang M, Song S, Dong K, Chen X, Liu X, Rouzi M, et al. Skin transcriptome reveals the intrinsic molecular mechanisms underlying hair follicle cycling in cashmere goats under natural and shortened photoperiod conditions. *Sci Rep*. (2017) 7:13502. doi: 10.1038/s41598-017-13986-w
 36. Geng R, Yuan C, Chen Y. Exploring differentially expressed genes by RNA-Seq in cashmere goat (*Capra hircus*) skin during hair follicle development and cycling. *PLoS One*. (2013) 8:e62704. doi: 10.1371/journal.pone.0062704
 37. Han W, Li X, Wang L, Wang H, Yang K, Wang Z, et al. Expression of fox-related genes in the skin follicles of Inner Mongolia cashmere goat. *Asian Australas J Anim Sci*. (2018) 31:316–26. doi: 10.5713/ajas.17.0115
 38. Hanley K, Kömüves LG, Bass NM, He SS, Jiang Y, Crumrine D, et al. Fetal epidermal differentiation and barrier development in vivo is accelerated by nuclear hormone receptor activators. *J Invest Dermatol*. (1999) 113:788–95. doi: 10.1046/j.1523-1747.1999.00743.x
 39. Hanley K, Jiang Y, Crumrine D, Bass NM, Appel R, Elias PM, et al. Activators of the nuclear hormone receptors PPARalpha and FXR accelerate the development of the fetal epidermal permeability barrier. *J Clin Invest*. (1997) 100:705–12. doi: 10.1172/jci119583
 40. Li HM, Liu P, Zhang XJ, Li LM, Jiang HY, Yan H, et al. Combined proteomics and transcriptomics reveal the genetic basis underlying the differentiation of skin appendages and immunity in pangolin. *Sci Rep*. (2020) 10:14566. doi: 10.1038/s41598-020-71513-w
 41. Nusse R, Clevers H. Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. *Cell*. (2017) 169:985–99. doi: 10.1016/j.cell.2017.05.016
 42. Shi Y, Shu B, Yang R, Xu Y, Xing B, Liu J, et al. Wnt and Notch signaling pathway involved in wound healing by targeting c-Myc and Hes1 separately. *Stem Cell Res Ther*. (2015) 6:120. doi: 10.1186/s13287-015-0103-4
 43. Vauclair S, Nicolas M, Barrandon Y, Radtke F. Notch1 is essential for postnatal hair follicle development and homeostasis. *Dev Biol*. (2005) 284:184–93. doi: 10.1016/j.ydbio.2005.05.018
 44. Mok KW, Saxena N, Heitman N, Grisanti L, Srivastava D, Muraro MJ, et al. Dermal condensate niche fate specification occurs prior to formation and is placode progenitor dependent. *Dev Cell*. (2019) 48:e35:32–48.e5. doi: 10.1016/j.devcel.2018.11.034
 45. Huelsken J, Vogel R, Erdmann B, Cotsarelis G, Birchmeier W. beta-catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell*. (2001) 105:533–45. doi: 10.1016/s0092-8674(01)00336-1
 46. Zhang Y, Tomann P, Andl T, Gallant NM, Huelsken J, Jerchow B, et al. Reciprocal requirements for EDA/EDAR/NF-kappaB and Wnt/beta-catenin signaling pathways in hair follicle induction. *Dev Cell*. (2009) 17:49–61. doi: 10.1016/j.devcel.2009.05.011
 47. Ge W, Tan SJ, Wang SH, Li L, Sun XF, Shen W, et al. Single-cell transcriptome profiling reveals dermal and epithelial cell fate decisions during embryonic hair follicle development. *Theranostics*. (2020) 10:7581–98. doi: 10.7150/thno.44306
 48. Greco V, Chen T, Rendl M, Schober M, Pasolli HA, Stokes N, et al. A two-step mechanism for stem cell activation during hair regeneration. *Cell Stem Cell*. (2009) 4:155–69. doi: 10.1016/j.stem.2008.12.009
 49. Yamane M, Seo J, Zhou Y, Asaba T, Tu S, Nanmo A, et al. Effects of the PI3K/Akt signaling pathway on the hair inductivity of human dermal papilla cells in hair beads. *J Biosci Bioeng*. (2022) 134:55–61. doi: 10.1016/j.jbiosc.2022.03.010
 50. Chen Y, Fan Z, Wang X, Mo M, Zeng SB, Xu RH, et al. PI3K/Akt signaling pathway is essential for de novo hair follicle regeneration. *Stem Cell Res Ther*. (2020) 11:144. doi: 10.1186/s13287-020-01650-6
 51. Liu G, Li S, Liu H, Zhu Y, Bai L, Sun H, et al. The functions of ocu-miR-205 in regulating hair follicle development in rex rabbits. *BMC Dev Biol*. (2020) 20:8. doi: 10.1186/s12861-020-00213-5

52. Hochmann J, Sobrinho JS, Villa LL, Sichero L. The Asian-American variant of human papillomavirus type 16 exhibits higher activation of MAPK and PI3K/AKT signaling pathways, transformation, migration and invasion of primary human keratinocytes. *Virology*. (2016) 492:145–54. doi: 10.1016/j.virol.2016.02.015
53. Hughes MW, Jiang TX, Plikus MV, Guerrero-Juarez CF, Lin CH, Schafer C, et al. Mx2 supports epidermal competency during wound-induced hair follicle Neogenesis. *J Invest Dermatol*. (2018) 138:2041–50. doi: 10.1016/j.jid.2018.02.043
54. Wang S, Luo Z, Zhang Y, Yuan D, Ge W, Wang X. The inconsistent regulation of HOXC13 on different keratins and the regulation mechanism on HOXC13 in cashmere goat (*Capra hircus*). *BMC Genomics*. (2018) 19:630. doi: 10.1186/s12864-018-5011-4
55. Huang D, Ding H, Wang Y, Cheng G, Wang X, Leng T, et al. Hair follicle transcriptome analysis reveals differentially expressed genes that regulate wool fiber diameter in angora rabbits. *Biology*. (2023) 12:12. doi: 10.3390/biology12030445
56. Mill P, Lee AW, Fukata Y, Tsutsumi R, Fukata M, Keighren M, et al. Palmitoylation regulates epidermal homeostasis and hair follicle differentiation. *PLoS Genet*. (2009) 5:e1000748. doi: 10.1371/journal.pgen.1000748
57. Chen Y, Bao Z, Liu M, Li J, Dai Y, Wang F, et al. Promoter methylation changes in KRT17: a novel epigenetic marker for wool production in angora rabbit. *Int J Mol Sci*. (2022) 23:23. doi: 10.3390/ijms23116077
58. Owens P, Bazzi H, Engelking E, Han G, Christiano AM, Wang XJ. Smad4-dependent desmoglein-4 expression contributes to hair follicle integrity. *Dev Biol*. (2008) 322:156–66. doi: 10.1016/j.ydbio.2008.07.020
59. Elias PM, Wakefield JS. Mechanisms of abnormal lamellar body secretion and the dysfunctional skin barrier in patients with atopic dermatitis. *J Allergy Clin Immunol*. (2014) 134:781–791.e1. doi: 10.1016/j.jaci.2014.05.048
60. Xu X, Mannik J, Kudryavtseva E, Lin KK, Flanagan LA, Spencer J, et al. Co-factors of LIM domains (Clims/Ldb/Nli) regulate corneal homeostasis and maintenance of hair follicle stem cells. *Dev Biol*. (2007) 312:484–500. doi: 10.1016/j.ydbio.2007.09.052
61. Lei M, Guo H, Qiu W, Lai X, Yang T, Widelitz RB, et al. Modulating hair follicle size with Wnt10b/DKK1 during hair regeneration. *Exp Dermatol*. (2014) 23:407–13. doi: 10.1111/exd.12416
62. Li H, Masieri FF, Schneider M, Bartella A, Gaus S, Hahnel S, et al. The middle part of the plucked hair follicle outer root sheath is identified as an area rich in lineage-specific stem cell markers. *Biomol Ther*. (2021) 11:11. doi: 10.3390/biom11020154
63. Wegner J, Loser K, Apsite G, Nischt R, Eckes B, Krieg T, et al. Laminin $\alpha 5$ in the keratinocyte basement membrane is required for epidermal-dermal intercommunication. *Matrix Biol: J Int Society for Matrix Biol*. (2016) 56:24–41. doi: 10.1016/j.matbio.2016.05.001
64. Qiao W, Li AG, Owens P, Xu X, Wang XJ, Deng CX. Hair follicle defects and squamous cell carcinoma formation in Smad4 conditional knockout mouse skin. *Oncogene*. (2006) 25:207–17. doi: 10.1038/sj.onc.1209029
65. Ohazama A, Johnson EB, Ota MS, Choi HY, Porntaveetus T, Oommen S, et al. Lrp4 modulates extracellular integration of cell signaling pathways in development. *PLoS One*. (2008) 3:e4092. doi: 10.1371/journal.pone.0004092
66. Ahn Y, Sims C, Logue JM, Weatherbee SD, Krumlauf R. Lrp4 and wise interplay controls the formation and patterning of mammary and other skin appendage placodes by modulating Wnt signaling. *Development*. (2013) 140:583–93. doi: 10.1242/dev.085118
67. Bao Q, Zhang X, Bao P, Liang C, Guo X, Yin M, et al. Genome-wide identification, characterization, and expression analysis of keratin genes (KRTs) family in yak (*Bos grunniens*). *Gene*. (2022) 818:146247. doi: 10.1016/j.gene.2022.146247
68. Xu Y, Wang S, Cao X, Yuan Z, Getachew T, Mwacharo JM, et al. The effect of EGFR1 on the proliferation of dermal papilla cells. *Gene*. (2022) 13:1242. doi: 10.3390/genes13071242
69. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. (2016) 534:652–7. doi: 10.1038/nature18606
70. Dussmann P, Pagel JL, Vogel S, Magnusson T, Zimmermann R, Wagner E, et al. Live in vivo imaging of Egr-1 promoter activity during neonatal development, liver regeneration and wound healing. *BMC Develop Biol*. (2011) 11:28. doi: 10.1186/1471-213x-11-28
71. Kwon HJ, Bhat N, Sweet EM, Cornell RA, Riley BB. Identification of early requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet*. (2010) 6:e1001133. doi: 10.1371/journal.pgen.1001133
72. Lv X, He M, Zhou H, Wang S, Cao X, Yuan Z, et al. SP1 and KROX20 regulate the proliferation of dermal papilla cells and target the CUX1 gene. *Animals: Open Access J MDPI*. (2024) 14:429. doi: 10.3390/ani14030429



OPEN ACCESS

EDITED BY

Fei Hao,
Northumbria University, United Kingdom

REVIEWED BY

Yingpeng Han,
Northeast Agricultural University, China
Kalpana Singh,
Guru Angad Dev Veterinary and Animal
Sciences University, India

*CORRESPONDENCE

Saeed Zerehdaran,
✉ zerehdaran@um.ac.ir

RECEIVED 09 April 2024

ACCEPTED 19 November 2024

PUBLISHED 19 December 2024

CITATION

Taheri S, Javadmanesh A and Zerehdaran S
(2024) Identification of selective sweep and
associated QTL traits in Iranian *Ovis aries* and
Ovis orientalis populations.
Front. Genet. 15:1414717.
doi: 10.3389/fgene.2024.1414717

COPYRIGHT

© 2024 Taheri, Javadmanesh and Zerehdaran.
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.

Identification of selective sweep and associated QTL traits in Iranian *Ovis aries* and *Ovis orientalis* populations

Sadegh Taheri, Ali Javadmanesh and Saeed Zerehdaran*

Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

Introduction: Identifying genomic regions under selection is the most challenging issue for improving important traits in animals. Few studies have focused on identifying genomic regions under selection in sheep. The aim of this study was to identify selective sweeps and to explore the relationship between these and quantitative trait loci (QTL) in both domestic and wild sheep species using single nucleotide polymorphism markers (SNPs).

Methods: Genomic data were obtained from the NextGen project, which included genotyping 20 domestic and 14 wild sheep using the Illumina Ovine SNP50K BeadChip. The XP-EHH, iHS, and RSB methods were employed to detect signatures of positive selection.

Results: The results of the iHS method indicated 405 and 275 selective sweeps in domestic and wild sheep, respectively. Additionally, RSB and XP-EHH analyses revealed approximately 398 and 479 selective sweeps in domestic and wild sheep, respectively. Some of the genes associated with important QTL traits in domestic sheep include *ADGRB3*, *CADM1*, *CAPN2*, *GALNT10*, *MTR*, *RELN*, and *USP25*, while in wild sheep, the relevant genes include *ACAN*, *ACO1*, *GADL1*, *MGST3*, and *PRDM16*. Selective sweeps identified in domestic sheep were associated with body weight, muscle weight, milk protein percentage, and milk yield. In contrast, selective sweeps found in wild sheep were linked to average daily gain, bone weight, carcass fat percentage, and dressing percentage.

Discussion: These results indicate that selection by humans and the environment have largely progressed in harmony, highlighting the importance of both economic and environmental traits for survival. Additionally, the identification of potential candidate genes associated with economic traits and genomic regions that have experienced selection can be utilized in sheep breeding programs. However, due to the incomplete information regarding the functional annotation of genes in sheep and the limited sample size, further research with a larger sample group is essential to gain a deeper understanding of the candidate genes linked to economic traits in both domestic and wild sheep. Advancing knowledge in this area can significantly enhance the effectiveness of breeding strategies. The quantitative trait loci identified in this study have the potential to be incorporated into breeding plans for both domestic and wild sheep.

KEYWORDS

domestic and wild sheep, iHS, RSB, selective sweep, QTL, XP-EHH

1 Introduction

Animal domestication has significantly influenced human life. Due to their high adaptability to various climates and diets, sheep were among the first animals to be domesticated (Zeder, 2008). Evidence suggests that Iran was a world center for farm animal domestication (Zeder, 1999). Archaeological findings indicate that sheep were domesticated approximately 9,000 years ago in what is now northern Iraq and parts of Iran (Kijas et al., 2009; Zeder, 1999). The wild sheep of Iran include three species: Vignei sheep (*Ovis vignei*), Asian mouflon (*O. orientalis*), and hybrids with varying chromosome numbers, which are found in the eastern and western regions and the Alborz Mountain range of northern Iran. Iranian wild sheep are recognized as the ancestors of domestic sheep worldwide, which may account for the diversity of domestic sheep breeds in Iran (Hamadallahmad et al., 2020). Domestic sheep are vital to Iran's agricultural economy and the livelihoods of its people, serving as a source of milk, meat, and other byproducts. To effectively manage the genetic resources of both wild and domestic animals, it is essential to understand the characteristics of different breeds. These include population size and structure, geographical distribution, optimal environmental conditions for production and performance, and levels of genetic diversity within and among breeds (Javadmanesh et al., 2022). Domestication has significantly altered the behavioral and morphological traits of these animals. In addition to domestication, selective breeding for increased production or specific behavioral and morphological traits has led to the emergence of highly divergent modern breeds (Flori et al., 2009). Post-domestication selection for economic and morphological traits can have lasting effects on the genomes of sheep. This selection, combined with natural adaptation to local environments, has resulted in the development of over a thousand distinct breeds of sheep (Kijas et al., 2012). These genetic features, along with the extensive genomic information related to economic traits, offer an opportunity to identify loci that are under selection (Hayes et al., 2008). Selection can lead to specific changes in both selected and neutral sites, with selection signatures—genomic footprints left by selection—used to identify these target areas (Kreitman, 2000). The recent availability of genomic data from domestic animals, coupled with advances in statistical tools, has made it possible to detect these signatures in specific species (Dong et al., 2013). Identifying selection signatures is a key focus for evolutionary geneticists, as it provides insight into the evolutionary processes that shape genomes and translates genomic data into functional information about important regions (Schlötterer, 2003). Several statistical methods are available for the investigation of selection signatures, including the integrated haplotype homozygosity score (iHS) (Voight et al., 2006), cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007), and the ratio of extended haplotype homozygosity between populations (RSB) (Tang et al., 2007). Advances in livestock genomics research have led to the recent development of QTL databases linked to the latest reference genomes of various livestock species, such as sheep, cows, pigs, and fish (Hui et al., 2021). These databases provide precise locations of genes and genomic regions associated with known traits in livestock species. As a result, researchers can use these

fine-mapped QTL traits to identify selection signatures related to desired traits in livestock species, which could be useful for breeding programs (Zeraatpisheh et al., 2023).

The current study aimed to detect the regions under natural and artificial selection in two wild and domestic Iranian sheep species using iHS, XP-EHH, and RSB methods. The communication between selective sweep and QTL traits was also identified using single nucleotide markers (SNP).

2 Materials and methods

2.1 Sampling and genotyping

The data utilized in this research were retrieved from the NextGen project (<http://projects.ensembl.org/nextgen/>). The objective of this project is to assess the genetic diversity of some domestic and wild species such as sheep, goat, and cattle by utilizing whole-genome sequencing and genotyping by microarray. For this particular study, samples were collected from 20 Iranian domestic (*O. aries*) and 14 Iranian wild sheep (*O. orientalis*). Samples were taken according to the Helsinki Declaration of 1975 (as revised in 2008) concerning animal rights, and this study was approved by the Animal Ethical Committee (3/51431), Ferdowsi University of Mashad, Mashhad, Iran (Taheri et al., 2022). Ear tissue samples were used for DNA extraction using a Macherey Nagel NucleoSpin 96 Tissue kit, following the manufacturer's instructions, as per Alberto et al. (2018). To integrate the whole-genome sequencing data with the variations in each animal's entire genome, the NextGen project's equivalent of the SNP50 BeadChip array was obtained from every individual utilizing PLINK v1.9 software.

2.2 Quality check and PCA analyses

To ensure the quality of the data, the PLINK version 1.9 application (Purcell et al., 2007) was used to perform quality control for the two populations. Individuals with a genotyping call rate of less than 99% were removed. Additionally, SNPs with a genotyping call rate of less than 99%, a minor allele frequency (MAF) of less than 5%, and a significant deviation from Hardy-Weinberg equilibrium (HWE) ($p < 10^{-6}$) were excluded (Taheri et al., 2023). Following this filtering process, the genotype data for both populations were combined, and common SNPs between them were selected.

To investigate the structures of the domestic and wild sheep populations and to identify animals that did not fit into their respective populations, principal component analysis (PCA) was conducted using a relative genomic matrix. PLINK version 1.9 was employed to perform this analysis.

2.3 Detection of selection sweep and significant SNPs

The three methods used in this study—iHS, RSB, and XP-EHH—are basically the same, based on the concept of extended

haplotype homozygosity (EHH), to increase the accuracy in identifying selection signatures. iHS is designed for use on a specific population, while XP-EHH and RSB are for comparing differences in selection between two populations. Three statistics were calculated using the EHH concept developed by Sabeti et al. (2002).

iHS analysis is a statistical method for identifying haplotype-based selection sweeps which involves the calculation of EHH from both ancestral (iHHA) and derived (iHHD) biallelic SNPs. iHS is a within-population value based on the rate of LD decay at a certain variant and is calculated as per Voight et al. (2006):

$$iHS = \ln \left(\frac{iHHA}{iHHD} \right).$$

The “rehh” package (Gautier and Vitalis, 2012) in R was utilized to separately estimate iHS values for SNPs with MAF ≥ 0.05 within domestic and wild populations. Large positive or negative values of iHS show long haplotypes harboring ancestral or derived alleles, respectively.

RSB is also calculated based on EHH. However, unlike iHS, RSB involves comparing the EHH patterns of the common alleles (as “iES”) between the two populations. Tang et al. (2007) defined RSB as:

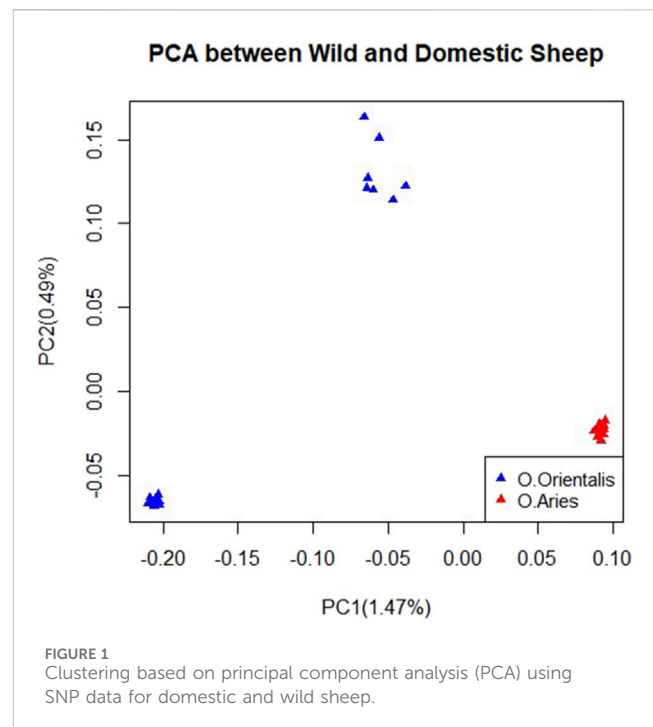
$$\ln(RSB) = \ln \left(\frac{iESpop1}{iESPOP2} \right).$$

To estimate RSB values, various iES statistics were estimated using the “rehh” package during the iHS analyses in each of the domestic and wild populations, as described previously. The RSB values were normalized in the R environment. Since the RSB value is directional, positive values indicate selection in the wild population, while negative values indicate selection in the domestic population.

XP-EHH was also utilized to identify any recent selection sweeps within the sheep genome. This method is based on evaluating LD decay across the genome and EHH values. For biallelic SNPs with alleles A and a, EHH was calculated using the following formula (Sabeti et al., 2002):

$$EHH = \frac{\sum_{i=1}^{h_x} h_x(ni^2)}{(na^2)(nA^2)},$$

where nA and na represent haplotype counts with alleles A and a, respectively, ni is the count for the i^{th} haplotype within a sub-population, and h_x represents the number of distinct haplotypes in a genomic area up to a distance x from the core locus. All SNPs located within one Mbp interval up- and downstream of a given SNP were taken into account, and EHH was integrated within these intervals for two groups. XP-EHH was estimated according to Sabeti et al. (2007). The highest negative and positive values indicate locations under selection in domestic and wild sheep populations, respectively. The “rehh” software was utilized to estimate XP-EHH, and the top 5% of estimated values were selected for further analysis using both domestic and wild populations (Tao et al., 2020). To increase the power of detecting selection signatures, we selected the overlapping areas between RSB, XP-EHH, and iHS techniques in domestic and wild sheep.



After identifying selection signatures using the three methods, similar locations in both populations were identified separately. The candidate genes related to these common locations were then selected using PLINK V1.9 (Purcell et al., 2007), utilizing a gene list provided from the sheep reference genome (<http://ncbi.nlm.nih.gov/genome/?term=sheep>).

2.4 Gene ontology and KEGG pathway analyses of genes related to significant SNPs

To identify gene ontology terms and significant metabolic KEGG pathways associated with genes related to significant SNPs, ClueGo version 2.5.6 was used, which is a Cytoscape plug-in that provides biological interpretations of genes (Bindea et al., 2009). Significant genes were then subjected to functional enrichment analysis using the Cytoscape application (Saito et al., 2012) and ClueGO 2.5.6 plug-in (Bindea et al., 2009). Official Gene Symbol was used as the input parameter, and *O. aries* was selected as the background organism. After Bonferroni correction for multiple testing, p -values < 0.05 were considered statistically significant.

2.5 Extraction of QTL traits associated with genes related to significant QTLs

Hu et al. (2022) utilized the AnimalQTL database (www.animalgenome.org/cgi-bin/QTLdb/OA/index) to locate sheep QTL traits previously reported. They then compared the genes identified in regions of high three methods with these QTL traits to determine whether any traits were under selection in Iranian domestic and wild sheep.

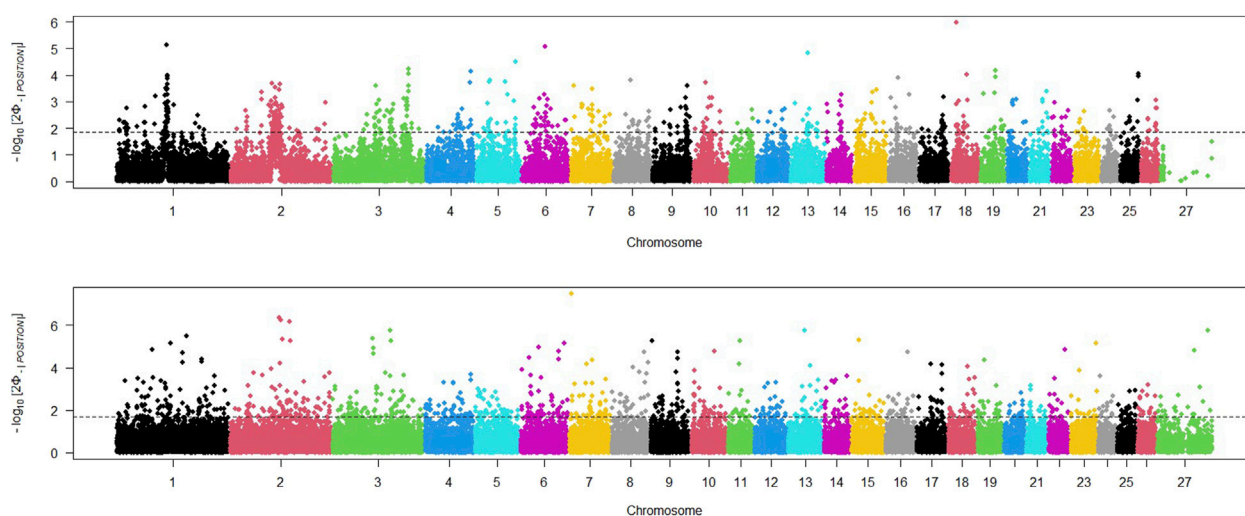


FIGURE 2
Distribution of iHS values across the genome in wild (above) and domestic (below) sheep; threshold indicates the top 5% SNPs with the largest iHS for each population.

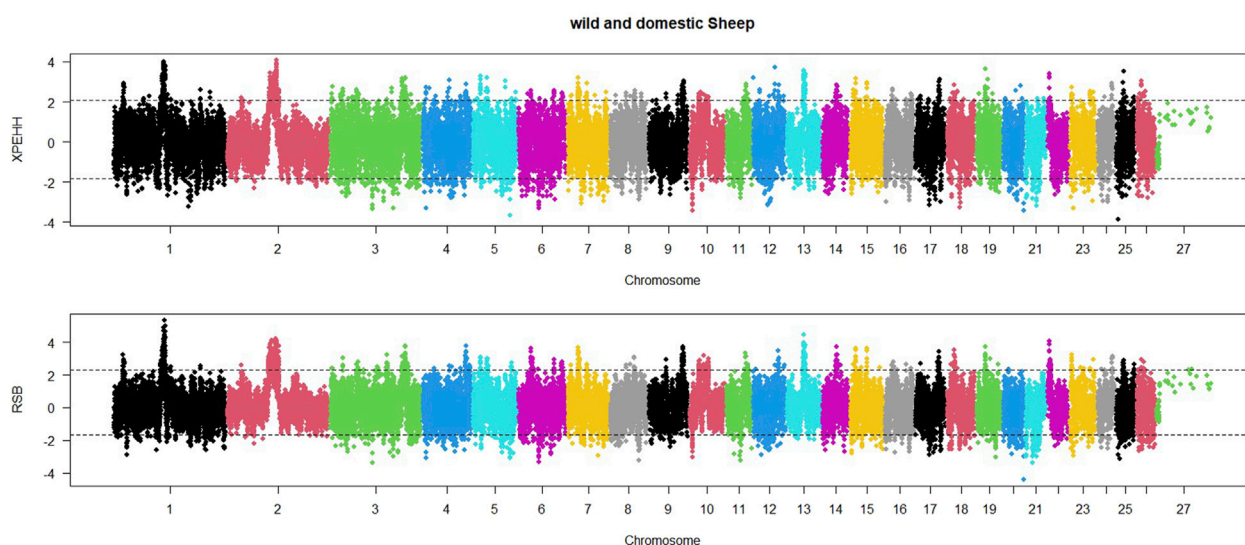


FIGURE 3
Distribution of XP-EHH and RSB across the genome between domestic (lower than 0) and wild (higher than 0) sheep. Thresholds indicate the top 5% SNPs for each population.

3 Results

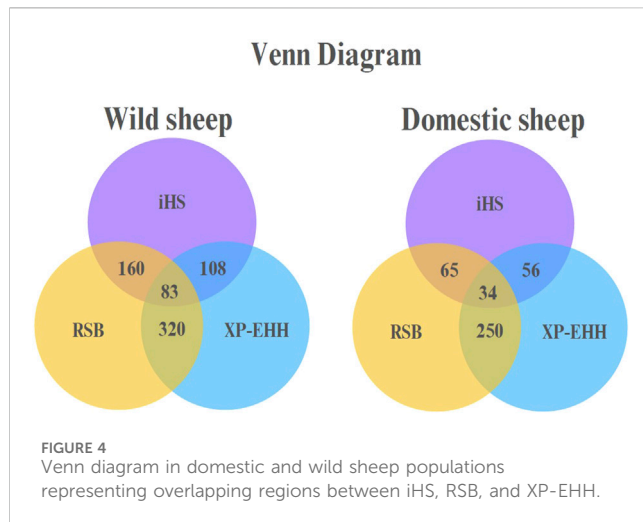
3.1 Quality control and PCA analyses

After performing quality control on the genomic data, we selected 33,167 and 35,479 SNPs for further analysis in domestic and wild sheep, respectively. To analyze the genetic connection between domestic and wild sheep and characterize their divergence, we conducted principal component analysis (PCA). The results of the PCA showed that domestic and wild sheep were clearly divided as shown in the first and second principal component factors. Furthermore, the results showed that the samples of domestic

sheep had lower within-population variation, while the wild sheep had more within-population variation, indicating two possible subpopulations (Figure 1).

3.2 Detection of selection sweep, significant SNPs, and their related genes

The study utilized three different statistical tests (iHS, RSB, XP-EHH) to detect any regions of the genome that had potentially undergone recent selection. The iHS test was applied to SNPs across the genome of both domestic and wild sheep populations, and a



Manhattan plot was generated to display the distribution of iHS signals across all chromosomes (Figure 2). We adjusted the threshold value to identify the top 5% of SNPs with the highest iHS for each population. The Manhattan plot for the wild sheep population indicated 275 locations with the highest iHS values ($iHS > 1.91$) on chromosomes 13, 18, 6, and 1. The Manhattan plot for the domestic sheep population revealed 405 regions with the highest signals ($iHS > 1.86$) on chromosomes 7, 2, and 3.

In order to identify selective sweep in the wild and domestic populations, an EHH-based RSB method was also utilized. The distribution of RSB values across sheep autosomal chromosomes is shown in Figure 3. Based on regions with RSB values lower than -1.92 , 421 regions were found to be under selection in the domestic sheep population. The lowest RSB values were observed on chromosomes 20, 11, and 6. Similarly, significant positive RSB values ($RSB > 2.05$) were identified in 376 regions in the wild sheep population. The highest RSB values were found on chromosomes 1, 2, and 13.

Additionally, a Manhattan plot for XP-EHH values within the sheep genome was generated (Figure 3). Based on regions with XP-EHH values lower than -2.04 , 467 regions were identified as under selection in the domestic sheep population. The most negative XP-EHH values were observed on chromosomes 25, 5, and 10. Similarly, according to significant selection signatures for XP-EHH values higher than 1.97, 490 regions were detected in the wild sheep population. The highest XP-EHH values were found on chromosomes 1, 2, and 12.

A Venn diagram for the domestic and wild sheep populations indicated that 34 genes in the domestic sheep and 83 genes in the wild sheep were shared by all three methods (Figure 4).

3.3 Gene ontology and KEGG pathway analyses of genes related to significant SNPs and associated QTL traits

We discovered that overlapping genes between the iHS, RSB, and XP-EHH methods were related to various processes in domestic (cell adhesion molecules, calcium-dependent cysteine-type endopeptidase activity, and cytoplasm) and wild (metabolic

pathways, nucleotide metabolism, and glutamatergic synapse) sheep (Table 1).

The quantitative trait locus report indicated that significant genes identified in the domestic sheep were connected to crucial traits such as body weight, bone density, bone weight, carcass fat percentage, and muscle weight. Additionally, candidate genes identified in the wild sheep were linked to average daily gain, milk fat percentage, milk yield, and muscle weight (Table 2) (more details in Supplementary Tables 1, 2).

4 Discussion

The results of the PCA indicated a clear distinction between the wild and domestic sheep populations (Figure 1). The use of genotypic relationship-based PCA has been a valuable tool in various studies to understand population structure and genetic connections between individuals (Sabeti et al., 2007). Domestic samples were collected from different regions, while wild samples were collected from the same region (Taheri et al., 2022). Our analysis revealed a more compact structure within the domestic sheep than the wild sheep, possibly due to intense selection pressure favoring economically desirable traits in domestic sheep (Taheri et al., 2022). However, it is important to note that the limited sample size in this study may lead to an underestimation of the true extent of diversity.

This study used three different methods (iHS, RSB, and XP-EHH) to improve the accuracy of identifying selective sweeps. The wild sheep population showed 135 locations, with the highest iHS values on chromosomes 1, 6, 13, and 18. In contrast, the domestic sheep population had 205 regions, with the strongest signals on chromosomes 2, 3, and 7. Using XP-EHH and FST, 93 candidate genomic regions were identified as harboring putative selective sweeps by Manzari et al. (2019) in three Iranian sheep breeds. The identified signatures of selection were related to multiple candidate genes involved in skeletal system, energy metabolism, growth, reproduction, and immune and nervous system traits. Eydivandi et al. (2021) applied FST, xp-EHH, Rsb, and FLK tests to identify selective sweeps. Their findings revealed 128, 207, 222, and 252 genomic regions, respectively, as candidates for selective sweeps. Additionally, nine overlapping candidate genes linked to disease resistance and climate adaptation were detected by all four tests. Alipanah et al. (2024) found the highest iHS coefficients under natural selection on chromosomes 3 and 2. Additionally, the XP-EHH results revealed that the highest XP-EHH coefficients under natural selection in European wild sheep, compared to Sardinian wild sheep, were observed on chromosome 3, while the reverse was true for Sardinian wild sheep compared to European wild sheep, with significant findings on chromosome 16.

By analyzing shared regions in the domestic population, we identified the genes *ADGRB3*, *CADMI*, *CAPN2*, *COPG2*, *DLC1*, *GALNT10*, *HIP1R*, *MTR*, *NEGR1*, *NRCAM*, *PML*, *PTPRM*, *RELN*, *RTN1*, *SDCCAG8*, and *USP25*. These are associated with economic traits such as longissimus muscle area, muscle weight, fecal egg count, milk yield, lean meat yield percentage, carcass fat percentage, bone density, udder attachment, meat palmitoleic acid content, somatic cell score, staple length, useful yield content, carcass bone percentage, meat polyunsaturated fatty acid content, bone weight, and hot carcass weight. *ADGRB3* has been identified as possibly related to fertilization and litter size by regulating oocyte

TABLE 1 GO terms and pathways along with the *p*-value of genes related to significant SNPs in domestic and wild sheep populations.

Category	Term	<i>p</i> -value	Gene
Wild sheep			
CC	Cytosol	0.006	<i>ZNF395, RALGAPA2, COPB1, RAP1GDS1, UAP1, ELP3, C9ORF72, PRDM16, UBR5, SPAG1, NEK1, TAF6, ACO1, RGS22, DCAF13, SH3GL1, and PLIN5</i>
BP	Positive regulation of GTPase activity	0.007	<i>RALGAPA2, RAP1GDS1, NF1, and C9ORF72</i>
CC	Presynapse	0.024	<i>KCNJ10, C9ORF72, and SH3GL1</i>
KEGG	Nucleotide metabolism	0.028	<i>UCK2, ENTPD6, and AK8</i>
CC	Glutamatergic synapse	0.037	<i>NAPB, ACAN, RNF19A, and SH3GL1</i>
BP	Regulation of phosphoprotein phosphatase activity	0.046	<i>TIPRL and TSC1</i>
KEGG	Metabolic pathways	0.050	<i>UCK2, PAH, MGST3, PRDM16, ENTPD6, GADL1, ACO1, UAP1, GALNTL6, and AK8</i>
Domestic sheep			
MF	Calcium-dependent cysteine-type endopeptidase activity	0.004	<i>CAPN13, CAPN8, and CAPN2</i>
CC	Cytoplasm	0.005	<i>TENM4, KCNIP1, RALGAPA1, SPATS2, RANBP17, MTR, CHD3, ARHGAP44, LOC101121185, RELN, XPOT, ELMO1, ARHGEF3, CAPN2, BBOF1, CPNE2, SDCCAG8, EIF4G1, and TRIM67</i>
CC	Golgi membrane	0.022	<i>GALNT14, LOC101112819, RTN1, COPG2, and GALNT10</i>
BP	Activation of cysteine-type endopeptidase activity involved in apoptotic process	0.023	<i>DLC1, HIP1R, and PML</i>
KEGG	Other types of O-glycan biosynthesis	0.024	<i>GALNT14, GXYLT1, and GALNT10</i>
BP	O-glycan processing	0.036	<i>GXYLT1 and GALNT10</i>
MF	SH3 domain binding	0.038	<i>ELMO1, HIP1R, and HCLS1</i>
BP	Negative regulation of protein ubiquitination involved in ubiquitin-dependent protein catabolic process	0.042	<i>HFE and PML</i>
BP	Negative regulation of angiogenesis	0.046	<i>ADGRB3, PTPRM, and PML</i>
MF	GTPase activator activity	0.048	<i>ARHGAP44, LOC101109993, RALGAPA1, and DLC1</i>
KEGG	Cell adhesion molecules	0.049	<i>CADM1, NEGRI, PTPRM, and NRCAM</i>
MF	SUMO binding	0.050	<i>USP25 and PML</i>

CC, cellular component, BP, biological process, MF, molecular function.

development in Hu sheep (Tao et al., 2021a). In a GWAS study, two significant SNPs within introns of this gene were found to be associated with a reduced number of parasite eggs in feces (Becker et al., 2022). *ADGRB3*, which is linked to the positive regulation of synapse assembly, has been identified in prolific sheep (Hernández-Montiel et al., 2022). *CADM* plays a crucial role in regulating embryo growth, body weight, fat metabolism, and energy balance. It is believed that this gene is involved in determining variations in body size (Xu et al., 2021). *CAPN2* is expressed widely in skeletal muscle, and previous research has linked it with gene expression and meat tenderness in various species (Knight et al., 2012). Ratzka et al. (2008) have shown that the *CAPN2* protein is essential for the normal growth of the preimplantation murine embryo.

COPG2 was identified in a study of receptor genes related to the progression of Johne’s disease in inbred sheep (Taylor et al., 2008). Additionally, it was found to be more active in the sheep lung and

brain on the maternal side prenatally (Duan et al., 2018). Research in cattle revealed that *COPG2* is expressed by both copies during fetal tissue development (Khatib, 2005).

Niciura et al. (2022) identified a link between *GALNT10* and weight traits. Similarly, Al Kalaldehy et al. (2019) found that *GALNT10*, which is responsible for producing a sugar molecule known as mucin-type O-glycan, is associated with the ability of sheep to resist gastrointestinal parasites. Jacobs et al. (2020) revealed that *HIP1R* is crucial for dendrite growth in sheep brain cells. *MTR* contains instructions for producing an enzyme known as 5-methyltetrahydrofolate-homocysteine methyltransferase, which plays a key role in converting homocysteine into methionine—an essential building block for proteins in the body (Zhang et al., 2012). Research suggests that *MTR* influences both the quantity and quality of wool produced by sheep (Rong et al., 2015). *NEGRI* is a cell adhesion molecule that belongs to the LON family of immunoglobulins, which also includes other molecules such as

TABLE 2 QTL traits associated with genes related to significant SNPs along with their genomic locations in domestic and wild populations.

CHR	Gene	SNP	SNP position	QTL
Domestic sheep				
1	NEGR1	ss1113654939	50204860	BONE_WT, BONEP, FA-C20:5, FA-C22:5, FATP, LMYP, MUSWT, and PUFA
1	USP25	ss1116150477 ss1116154441	152500264 152601830	BONE_WT, BONEP, FA-C20:5, FA-C22:5, FATP, FLYD, LMYP, MUSWT, and PUFA
4	COPG2	ss1129997095 ss1129997769	103086813 103151963	CVFD_PRI
4	NRCAM	ss1129534548	55685078	CVFD_PRI
4	RELN	ss1129484817 ss1129485872	50358993 50458109	CVFD_PRI
5	GALNT10	ss1130834348	68288153	FA-C16:1
7	RTN1	ss1133283846	74925794	CVFD_PRI, SL
9	ADGRB3	ss1134565186 ss1134567992	5336181 5535614	LMA and MUSWT
12	CAPN2	ss1137368707	29531925	BDENS, FATP, LMYP, and MY
12	SDCCAG8	ss1137447517	36847011	BDENS, FATP, LMYP, and MY
15	CADM1	ss1139557872	28497013	FECGEN
17	HIP1R	ss1141565009	61655059	BDENS and SCS
18	PML	ss1141994265	32479250	FA-C20:1, FATP, SL, and TESTWT
23	PTPRM	ss1145103439	45903918	FATP, FATWT, HCWT, LMYP, MFY_180D, and MY
25	MTR	ss1145839414	9173663	CVFD_PRI, MFDIAM, MFPER, SL, TESTWT, and UYC
26	DLC1	ss1146742924 ss1146743820 ss1146744870	25957009 26043393	MUSWT, Stature, UDDATT, and WORMCT
Wild sheep				
1	KCNJ10	ss1115230121	117944910	BDENS, BFLUMB3, BONE_WT, FA-C20:5, FA-C22:5, FATP, FECGEN, LMYP, MDLUMB3, MUSWT, and PUFA
1	MGST3	ss1115375784	125162941	BDENS, BFLUMB3, BONE_WT, FA-C20:5, FA-C22:5, FATP, FECGEN, LMYP, MDLUMB3, MUSWT, and PUFA
2	ACO1	ss1121421351 ss1121422525	109766460 109804281	FA-C18:3, FA-C20:4, FA-C20:5, FA-C22:5, HCWT, LATRICH_2, MFPER, and SCS
2	GALNTL6	ss1121533683 ss1121534742 ss1121534985 ss1121536161 ss1121537352 ss1121537767 ss1121538473 ss1121540364 ss1121541584 ss1121544347 ss1121548053 ss1121548153 ss1121548799 ss1121549287 ss1121549759 ss1121554393 ss1121568385 ss1121570596 ss1121571444 ss1121573654	116265153 116350674 116362603 116417246 116465257 116500683 116534114 116594826 116635922 116763390 116898149 116905792 116953479 116992107 117072252 117115355 117485350 117573048 117612053 117702396	FA-C18:3, FA-C20:4, FA-C20:5, FA-C22:5, HCWT, LATRICH_2, MFPER, and NFEC

(Continued on following page)

TABLE 2 (Continued) QTL traits associated with genes related to significant SNPs along with their genomic locations in domestic and wild populations.

CHR	Gene	SNP	SNP position	QTL
		ss1121577773 ss1121580557	117736662 117849339	
2	ZNF395	ss1158590803	111016299	FA-C18:3, FA-C20:4, FA-C20:5, FA-C22:5, HCWT, LATRICH_2, MFPER, and SCS
3	AK8	ss1124538240	4201867	HFEC and SL
3	TSC1	ss1124537252	4132992	HFEC and SL
5	PLIN5	ss1130359135	18089207	FA-C16:1
9	SPAG1	ss1135395144 ss1135395539	84948639 84978362	HCWT, LMA, MFY_180D, and MUSWT
11	NF1	ss1209682579 ss1209684264	44753284 44896884	HCWT, JAWL, and LATRICH_2
12	PRDM16	ss1137575478 ss1137575655	53644681 53663779	FATP and LMYP
13	RALGAPA2	ss1138263148 ss1138263481 ss1138264060 ss1138264441 ss1138264932 ss1138265074 ss1138265616	41284446 41314389 41371630 41394956 41434517 41443275 41483549	MUSWT and SAOS
18	ACAN	ss1218825229	17861995	FA-C20:1, MY, SL, and TESTWT
19	GADL1	ss1219489564	5622653	ASREP and DRESSING

AMDG, age at maximum daily gain; ADG, average daily gain; BFLUMB3, back fat at third lumbar; BW, body weight; BDENS, bone density; BONE_WT, carcass bone weight; BONEP, carcass bone percentage; FATP, carcass fat percentage; DRESSING, dressing percentage; FATWT, carcass fat weight; FECGEN, fecal egg count; FCURV, fiber curvature; FLYD, fleece yield; HFEC, *Haemonchus contortus* FEC; HO, horns; HCWT, hot carcass weight; IGA, immunoglobulin A level; IGG, immunoglobulin G level; IOA, inherited ovine arthrogryposis; INTFAT, internal fat amount; JAWL, jaw length; LMYP, lean meat yield percentage; LMA, longissimus muscle area; MFDIAM, mean fiber diameter; FA-C20:4, meat arachidonic acid content; FA-C18:1, meat cis-vaccenic acid content; FA-C22:5, meat docosapentaenoic acid content; FA-C20:5, meat eicosapentaenoic acid content; FA-C20:1, meat gadoleic acid content; FA-C18:2, meat linoleic acid content; FA-C18:3, meat linolenic acid content; FA-C14:0, meat myristic acid content; FA-C18:1, meat oleic acid content; FA-C16:0, meat palmitic acid content; FA-C16:1, meat palmitoleic acid content; PUFA, meat polyunsaturated fatty acid content; FA-C14:0, meat stearic acid content; MCARPL, metacarpal length; MFPER, milk fat percentage; MFY_180D, milk fat yield; MLACT, milk lactose yield; MPUFA, milk polyunsaturated fatty acid content; PP, milk protein percentage; PY, milk protein yield; MYPERS, milk yield persistency; MY, milk yield; MDLUMB3, muscle depth at third lumbar; MUSWT, muscle weight in carcass; NFEC, *Nematodirus* FEC; CVFD_PRI, primary fiber diameter coefficient of variance; RLEGS, rear leg set; ASREP, reproductive seasonality; SAOS, *Salmonella* Abortusovis susceptibility; SCS, somatic cell score; SL, staple length; Stature: stature; SCFA, subcutaneous fat area; SCFT, subcutaneous fat thickness; TESTWT, testes weight; TOTBONE, total bone; LATRICH_2, *Trichostrongylus* adult and larva count; TFEC_1, *Trichostrongylus colubriformis* FEC; UDDATT, udder attachment; UYC, useful yield content; WORMCT, worm count.

limbic system-associated membrane protein and neurotrimin (Noh et al., 2019). *NEGR1* has been associated with feed efficiency in beef cattle (Seabury et al., 2017) and somatic cell score in sheep (Mohammadi et al., 2022). Specifically, *NEGR1* is more prevalent in the cell adhesion molecular pathway, which plays a role in the body’s defense against disease in cattle (Liu et al., 2020). *NEGR1* is a protein that can be activated as needed and is involved in the growth, specialization, and death of cells in blood vessels (Wang et al., 2021). *NRCAM* is linked to cells in the nervous system of sheep and with molecules that aid cell adhesion (Alvarez-Franco et al., 2021). Recent studies have demonstrated that *NRCAM*, produced by uterine lining cells, can enhance the body’s response to progesterin hormones by altering gene behavior (Cheng et al., 2022). In ruminants, research has revealed that a specific gene called *PML* plays a crucial role in inhibiting tumor growth, indicating a potentially enhanced cancer prevention system in these animals (Wang et al., 2019). Increased expression of *PTPRM* in small intestinal neuroendocrine tumor cells reduces cell growth and division and induces cell death (Barazeghi et al., 2019). *RELN* has been observed to be more expressed in susceptible animals (McRae

et al., 2014). Comparing Suffolk and Texel sheep infected with *Teladorsagia circumcincta*, *RELN* was found to be more expressed in Suffolk sheep 3 days post-infection (Ahmed, 2013). Most notably, the expression of *RELN* is concentrated in the theca cells of dominant follicles, where it modulates downstream signaling pathways through paracrine interaction with *LRP8* in granulosa cells (Fayad et al., 2007), regulating the final stages of follicle growth (Nivet et al., 2013). Studies suggest an association between *RELN* and protein kinase activity, contributing to progestogenic pathways, while also highlighting a negative impact of MAP due to *RELN* suppression. *RELN*, a glycoprotein in the extracellular matrix, is involved in various cellular functions, including the MAPK pathway, which is vital for germinal vesicle breakdown and oocyte maturation (Yang et al., 2018). Aboul-Naga et al. (2022) uncovered crucial genetic variations associated with heat tolerance in the gene *RTN1*. *RTN1* is part of the reticulon family residing in the endoplasmic reticulum (ER), which is implicated in hormone release and membrane mobility in nerve cells and is closely linked to ER stress (Chiurchiu et al., 2014). In genome-wide analyses of goats and sheep, *SDCCAG8* was identified as under selection, affecting

reproduction and the TGF pathway, which governs the number of offspring in goats and sheep (Tao et al., 2021b). *SDCCAG8* is also involved in cellular component organization (Han M. et al., 2022). Zhang et al. (2022) confirmed that *USP25* is correlated with the annual reproductive cycle in sheep by investigating various sheep breeds with distinct characteristics.

Identifying common segments in the wild population allowed us to accurately locate the genes *ACAN*, *ACOI*, *AK8*, *GADL1*, *GALNTL6*, *KCNJ10*, *MGST3*, *NF1*, *PLIN5*, *PRDM16*, *RALGAP2*, *SPAG1*, *TSC1*, and *ZNF395*. These genes were associated with several economic traits in the wild population, including muscle weight, milk yield, lean meat percentage, carcass fat percentage, bone density, somatic cell score, staple length, meat polyunsaturated fatty acid content, hot carcass weight, testes weight, milk fat percentage, reproductive seasonality, and jaw length. *ACAN* encodes the aggrecan protein, which is a type of proteoglycan. Aggrecan is the most abundant proteoglycan found in cartilage, a tough and flexible tissue that forms a significant part of the skeleton during early development. The majority of cartilage eventually transforms into bone through a process called ossification, with the exception of the cartilage that remains to cover and protect the ends of bones, along with that found in the nose, airways, and external ears (Mancioppi et al., 2021). *ACAN* mutations can cause skeletal disorders such as osteochondrosis and skeletal dysplasia, which affect height (Sabeti-Aghabozorgi et al., 2022).

Manzari et al. (2019) investigated genes related to skeletal and tail growth, noting that genes such as *ACAN* show signs of selection. *GALNTL6* is an important gene for the ability of sheep to resist gastrointestinal parasites (Pratap et al., 2024). *MGST3* is associated with muscle tissue in sheep and is important for amino acid metabolism (Thameem et al., 2003). *NF1* plays a role in various cellular functions and is linked to sheep reproductive performance. It is crucial for increasing the number of lambs in Texel sheep and for adapting to high-altitude hypoxia (XU et al., 2018). *PLIN5* is involved in fat metabolism and insulin regulation (Zhang et al., 2022). *PRDM16* plays a role in fat cell formation and embryonic development (Chi and Cohen, 2016). *SPAG1* was found to influence fertility in mammals (Faraji et al., 2021). *TSC1* is linked to body size variability in sheep breeds (Cao et al., 2015), while *ZNF395* is associated with growth traits and the development of fat cells. It may also play a role in obesity and metabolic syndrome. *ZNF395* expression levels in Lanzhou fat-tailed sheep were notably higher than in small-tailed Han sheep (Erdenee et al., 2020).

ADAM9, *ARHGAP42*, and *CUTC* genes were identified as common genes between domestic and wild sheep species. The *ADAM9* gene plays a crucial role in the nervous, cardiovascular, and muscular systems. In the nervous system, *ADAM9* is involved in neuron migration, axon growth, and synapse formation (Cho, 2012). In the cardiovascular system, it contributes to the proliferation and migration of vascular endothelial cells (Chou et al., 2020). In muscle tissue, *ADAM9* is essential for muscle cell proliferation and differentiation (Ahmed et al., 2017). Furthermore, another study showed significant differences in *ADAM9* expression in sheep skin tissues with varying wool fineness (Tian et al., 2017). The *ARHGAP42* gene is associated with hair follicle development and wool production traits in sheep (Sun et al., 2021). This gene regulates angiogenesis in cattle (Diao et al., 2019). Additionally, this gene has been linked to hypertension in a study examining the interaction between genomics and diet in adult males (Imaizumi et al., 2017). The *CUTC* gene was evaluated in the methane

yield metagenome and metatranscriptome datasets of sheep (Kelly et al., 2019).

Quantitative trait locus analysis revealed that significant genes identified in domestic sheep populations were primarily linked to economically important traits, including body weight, bone structure, and muscle development. In contrast, candidate genes found in wild sheep populations were associated with structural and immune traits, such as reproductive behaviors and resistance to parasites, *Salmonella abortusovis*, and mastitis. Additionally, there were candidate genes in wild sheep related to performance traits, including body weight, muscle growth, milk yield, and milk fat. These results clearly show that selection for higher performance in domestic sheep makes them more sensitive to environmental stressors and diseases, while there is a balance between performance and biologically important traits in wild sheep population. Lv et al. (2021) demonstrated that domestic sheep may have acquired beneficial alleles of various immune and sensory genes through natural or managed hybridization with their wild sheep.

5 Conclusion

Our findings suggest that artificial selection by humans, with an emphasis on economically important traits in domestic sheep, has weakened the balance between economic and environmental traits. This balance appears to be crucial for the survival in wild sheep. By identifying potential candidate genes associated with economic and survival traits, along with their genomic regions that have undergone changes due to selection, these insights can be utilized in breeding programs for sheep. However, due to the incomplete information regarding the functional annotation of genes within sheep species and the limited sample size studied, further research with a larger number of samples is necessary to gain a deeper understanding of candidate genes for critical economic traits in both domestic and wild sheep. Advancing knowledge in this area can significantly enhance the design of effective breeding strategies.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at: <http://projects.ensembl.org/nextgen/>.

Ethics statement

The animal study was approved by the Animal Ethics Committee (3/51431), Ferdowsi University of Mashad. The study was conducted in accordance with local legislation and institutional requirements.

Author contributions

ST: data curation, investigation, methodology, software, and writing—original draft. AJ: methodology, supervision, and writing—review and editing. SZ: supervision, validation, writing—original draft, and writing—review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Ferdowsi University of Mashhad. This study makes use of data generated by the NextGen Consortium. The European Union's Seventh Framework Programme (FP7/2010-2014) provided funding for the project under grant agreement no 244356 - "NextGen".

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.

References

- Aboul-Naga, A. M., Alsamman, A. M., El Allali, A., Elshafie, M. H., Abdelsabour, T. H., Hamwih, A., et al. (2022). The genome-wide analysis identified candidate variants and genes associated with heat stress adaptation in Egyptian sheep breeds. *Front. Genet.* 13, 898522. doi:10.3389/fgene.2022.898522
- Ahmed, A. M. (2013). *Immunological and transcriptomic analysis of genetically resistant and susceptible sheep to gastrointestinal nematodes*. Dublin: University College Dublin. PhD Thesis.
- Ahmed, S. I., Ibrahim, M. E., and Khalil, E. A. (2017). High altitude and pre-clampsia: adaptation or protection. *Med. Hypotheses* 104, 128–132.
- Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., De Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* 9 (1), 813. doi:10.1038/s41467-018-03206-y
- Alipanah, M., Mazloom, S. M., and Gharari, F. (2024). Detection of selective sweep in European wild sheep breeds. *3Biotech* 14 (122), 122. doi:10.1007/s13205-024-03964-1
- Al Kalaladeh, M., Gibson, J., Lee, S. H., Gondro, C., and Van Der Werf, J. H. (2019). Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genet. Sel. Evol.* 51, 37–18. doi:10.1186/s12711-019-0479-1
- Alvarez-Franco, A., Rouco, R., Ramirez, R. J., Guerrero-Serna, G., Tiana, M., Cogliati, S., et al. (2021). Transcriptome and proteome mapping in the sheep atria reveal molecular features of atrial fibrillation progression. *Cardiovasc. Res.* 117 (7), 1760–1775. doi:10.1093/cvr/cvaa307
- Barazeghi, E., Hellman, P., Westin, G., and Ståhlberg, P. (2019). PTPRM, a candidate tumor suppressor gene in small intestinal neuroendocrine tumors. *Endocr. Connect.* 8 (8), 1126–1135. doi:10.1530/EC-19-0279
- Becker, G. M., Burke, J. M., Lewis, R. M., Rosen, B. D., Van Tassell, C. P., Notter, D. R., et al. (2022). Variants within genes EDIL3 and ADGRB3 are associated with divergent fecal egg counts in Katahdin sheep at weaning. *Front. Genet.* 13, 817319. doi:10.3389/fgene.2022.817319
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi:10.1093/bioinformatics/btp101
- Cao, J., Wei, C., Liu, D., Wang, H., Wu, M., Xie, Z., et al. (2015). DNA methylation Landscape of body size variation in sheep. *Sci. Rep.* 5 (1), 13950. doi:10.1038/srep13950
- Cheng, Y., Xie, L., Xu, Z., Hao, M., Yang, B., Shan, W., et al. (2022). NrCAM secreted by endometrial stromal cells enhances the progesterone sensitivity of endometrial cancer cells through epigenetic modulation of PRB. *Cancer Gene Ther.* 29 (10), 1452–1462. doi:10.1038/s41417-022-00467-0
- Chi, J., and Cohen, P. (2016). The multifaceted roles of PRDM16: adipose biology and beyond. *Trends Endocrinol. Metabolism* 27 (1), p11–p23. doi:10.1016/j.tem.2015.11.005
- Chiurciu, V., Maccarrone, M., and Orlacchio, A. (2014). The role of reticulons in neurodegenerative diseases. *Neuromol. Med.* 16, 3–15. doi:10.1007/s12017-013-8271-9
- Cho, C. (2012). Testicular and epididymal ADAMs: expression and function during fertilization. *Nat. Rev. Urol.* 9 (10), 550–560.
- Chou, C. W., Huang, Y. K., Kuo, T. T., Liu, J. P., and Sher, Y. P. (2020). An overview of ADAM9: structure, activation, and regulation in human diseases. *Int. J. Mol. Sci.* 21 (20), 7790.
- Diao, S., Huang, S., Xu, Z., Ye, S., Yuan, X., and Chen, Z. (2019). Genetic diversity of indigenous pigs from South China area revealed by SNP array. *Animals* 9 (6), 361.
- Dong, Y., Xie, M., Jiang, Y. U., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31 (2), 135–141. doi:10.1038/nbt.2478
- Duan, J., Zhang, M., Flock, K., Seesi, S. A., Mandoiu, I., Jones, A., et al. (2018). Effects of maternal nutrition on the expression of genomic imprinted genes in ovine fetuses. *Epigenetics* 13 (8), 793–807. doi:10.1080/15592294.2018.1503489
- Erdenee, S., Li, J., Kang, Z., Xu, H., Zang, R., Cao, X., et al. (2020). Sheep zinc finger proteins 395 (ZNF395): insertion/deletion variations, associations with growth traits, and mRNA expression. *Anim. Biotechnol.* 31 (3), 237–244. doi:10.1080/10495398.2019.1585865
- Eydivandi, S., Roudbar, M. A., Ardestani, S. S., Momen, M., and Sahana, G. (2021). A selection signatures study among Middle Eastern and European sheep breeds. *J. Anim. Breed. Genet.* 138 (5), 574–588. doi:10.1111/jbg.12536
- Faraji, S., Sharafi, M., Shahverdi, A. H., and Fathi, R. (2021). Sperm associated antigens: vigorous influencers in life. *Cell J.* 23 (5), 495–502. doi:10.22074/cellj.2021.7377
- Fayad, T., Lefebvre, R., Nimpf, J., Silversides, D. W., and Lussier, J. G. (2007). Low-density lipoprotein receptor-related protein 8 (LRP8) is upregulated in granulosa cells of bovine dominant follicle: molecular characterization and spatio-temporal expression studies. *Biol. Reprod.* 76 (3), 466–475. doi:10.1095/biolreprod.106.057216
- Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., et al. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS One* 4 (8), e6595. doi:10.1371/journal.pone.0006595
- Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28 (8), 1176–1177. doi:10.1093/bioinformatics/bts115
- Hamadallah, A., Almezad, M., and Javadmanesh, A. (2020). Genetic similarity comparison between some Iranian and Middle Eastern sheep breeds using mitochondrial control region sequencing. *Dyson. Life Sci.* 1 (1), 20–24. doi:10.30493/dls.2020.105085
- Han, M., Liang, C., Liu, Y., He, X., and Chu, M. (2022). Integrated transcriptome analysis reveals the crucial mRNAs and miRNAs related to fecundity in the hypothalamus of yunshang black goats during the luteal phase. *Animals* 12 (23), 3397. doi:10.3390/ani12233397
- Hayes, B. J., Lien, S., Nilsen, H., Olsen, H. G., Berg, P., MacEachern, S., et al. (2008). The origin of selection signatures on bovine chromosome 6. *Anim. Genet.* 39 (2), 105–111. doi:10.1111/j.1365-2052.2007.01683.x
- Hernández-Montiel, W., Cob-Calan, N. N., Cahuich-Tzuc, L. E., Rueda, J. A., Quiroz-Valiente, J., Meza-Villalva, V., et al. (2022). Runs of homozygosity and gene identification in pelibuey sheep using genomic data. *Diversity* 14 (7), 522. doi:10.3390/d14070522

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.2024.1414717/full#supplementary-material>

- Hu, Z. L., Park, C. A., and Reecy, J. M. (2022). Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic. acids. Res.* 50 (D1), D956–D961. doi:10.1093/nar/gkab1116
- Hui, T., Zheng, Y., Yue, C., Wang, Y., Bai, Z., Sun, J., et al. (2021). Screening of cashmere fineness-related genes and their ceRNA network construction in cashmere goats. *Sci. Rep.* 11 (1), 21977. doi:10.1038/s41598-021-01203-8
- Imaizumi, T., Ando, M., Nakatochi, M., Maruyama, S., Yasuda, Y., Honda, H., et al. (2017). Association of interactions between dietary salt consumption and hypertension-susceptibility genetic polymorphisms with blood pressure among Japanese male workers. *Clin. Exp. Nephrol.* 21, 457–464.
- Jacobs, A., Hagin, M., Shugol, M., Shomron, N., Pillar, N., Fañanás, L., et al. (2020). The black sheep of the family-whole-exome sequencing in family of lithium response discordant bipolar monozygotic twins. *Eur. Neuropsychopharmacol.* 34, 19–27. doi:10.1016/j.euroneuro.2020.03.009
- Javadmanesh, A., Ghovvati Rodsari, S., Soltani, M., and Nassiry, M. (2022). Genetic diversity of ural population in Northeast of Iran. *Iran. J. Anim. Biosyst.* 18 (2), 185–193. doi:10.22067/ijab.2022.74490.1024
- Kelly, W. J., Leahy, S. C., Kamke, J., Soni, P., Koike, S., and Mackie, R. (2019). Occurrence and expression of genes encoding methyl-compound production in rumen bacteria. *Anim. Microbiome* 1, 1–13.
- Khatib, H. (2005). The COPG2, DCN, and SDHD genes are biallelically expressed in cattle. *Mamm. Genome.* 16, 545–552. doi:10.1007/s00335-005-0029-8
- 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 (2), e1001258. doi:10.1371/journal.pbio.1001258
- Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., et al. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS. One.* 4 (3), e4668. doi:10.1371/journal.pone.0004668
- Knight, M. I., Daetwyler, H. D., Hayes, B. J., Hayden, M. J., Ball, A. J., Pethick, D. W., et al. (2012). Discovery and trait association of single nucleotide polymorphisms from gene regions of influence on meat tenderness and long-chain omega-3 fatty acid content in Australian lamb. *Anim. Prod. Sci.* 52 (7), 591–600. doi:10.1071/AN11229
- Kreitman, M. (2000). Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.* 1 (1), 539–559. doi:10.1146/annurev.genom.1.1.539
- Liu, Y., Xu, L., Yang, L., Li, J., Li, Y., Liu, D., et al. (2020). Discovery of genomic characteristics and selection signatures in Southern Chinese local cattle. *Front. Genet.* 11, 533052. doi:10.3389/fgene.2020.533052
- Lv, F. H., Cao, Y. H., Liu, G. -J., Luo, L. -Y., Lu, R., Liu, M. -J., et al. (2021). Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, a agronomically important loci. *Mol. Biol. Evol.* 39 (2), msab353. doi:10.1093/molbev/msab353
- Mancioppi, V., Prodam, F., Mellone, S., Ricotti, R., Giglione, E., Grasso, N., et al. (2021). Retrospective diagnosis of a novel ACAN pathogenic variant in a family with short stature: a case report and review of the literature. *Front. Genet.* 12, 708864. doi:10.3389/fgene.2021.708864
- Manzari, Z., Mehrabani-Yeganeh, H., Nejati-Javaremi, A., Moradi, M. H., and Gholizadeh, M. (2019). Detecting selection signatures in three Iranian sheep breeds. *Anim. Genet.* 50 (3), 298–302. doi:10.1111/age.12772
- McRae, K. M., McEwan, J. C., Dodds, K. G., and Gemmell, N. J. (2014). Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *Bmc. Genomics* 15, 637. doi:10.1186/1471-2164-15-637
- Mohammadi, H., Farahani, A. H. K., Moradi, M. H., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., et al. (2022). Weighted single-step genome-wide association study uncovers known and novel candidate genomic regions for milk production traits and somatic cell score in Valle del Belice dairy sheep. *Animals* 12 (9), 1155. doi:10.3390/ani12091155
- Niciura, S. C. M., Benavides, M. V., Okino, C. H., Ibelli, A. M. G., Minho, A. P., Esteves, S. N., et al. (2022). Genome-wide association study for *Haemonchus contortus* resistance in Morada Nova sheep. *Pathogens* 11 (8), 939. doi:10.3390/pathogens11080939
- Nivet, A. L., Vigneault, C., Blondin, P., and Sirard, M. A. (2013). Changes in granulosa cells' gene expression associated with increased oocyte competence in bovine. *Reproduction* 145 (6), 555–565. doi:10.1530/REP-13-0032
- Noh, K., Lee, H., Choi, T. Y., Joo, Y., Kim, S. J., Kim, H., et al. (2019). Negr1 controls adult hippocampal neurogenesis and affective behaviors. *Mol. Psychiatry* 24 (8), 1189–1205. doi:10.1038/s41380-018-0347-3
- Pratap, R., Chennuru, S., Krovvidi, S., Chitithoti, J., and Pentala, R. K. (2024). Putative SNPs in *Ovar-DRB1* and *GALNTL6* genes conferring susceptibility to natural infection of *haemonchus contortus* in Southern Indian sheep. *Acta Parasitol.* 69 (1), 583–590. doi:10.1007/s11686-023-00778-8
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795
- Ratzka, A., Kalus, I., Moser, M., Dierks, T., Mundlos, S., and Vortkamp, A. (2008). Redundant function of the heparan sulfate 6-O-endosulfatases Sulf1 and Sulf2 during skeletal development. *Dev. Dyn.* 237 (2), 339–353. doi:10.1002/dvdy.21423
- 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. Anim. Sci.* 93 (10), 4601–4609. doi:10.2527/jas.2015-8963
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419 (6909), 832–837. doi:10.1038/nature01140
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449 (7164), 913–918. doi:10.1038/nature06250
- Sabeti-Aghabozorgi, A., Ahangari, N., Mobini, M., Vakili, R., Doosti, M., Vakili, S., et al. (2022). A novel heterozygous ACAN variant in an Iranian family with short stature: a case report. *Int. J. Pediatr.* 10 (9), 16729–16737. doi:10.22038/ijp.2022.63946.4859
- Saito, R., Smoot, M. E., Ono, K., Ruschinski, J., Wang, P. L., Lotia, S., et al. (2012). A travel guide to Cytoscape plugins. *Nat. Methods* 9 (11), 1069–1076. doi:10.1038/nmeth.2212
- Schlötterer, C. (2003). Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends. Genet.* 19 (1), 32–38. doi:10.1016/S0168-9525(02)00012-4
- Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., et al. (2017). Genome-wide association study for feed efficiency and growth traits in US beef cattle. *Bmc. Genomics* 18, 386. doi:10.1186/s12864-017-3754-y
- Sun, W., Ma, S., and Ma, Y. (2021). On hair follicle development and wool production traits in sheep: a review. *Intl. J. Agric. Biol.* 25, 450–454. doi:10.17957/IJAB/15.1687
- Taheri, S., Saedi, N., Zerehdaran, S., and Javadmanesh, A. (2023). Identification of selection signatures in *Capra hircus* and *Capra aegagrus* in Iran. *Anim. Sci. J.* 94 (1), e13864. doi:10.1111/asj.13864
- Taheri, S., Zerehdaran, S., and Javadmanesh, A. (2022). Genetic diversity in some domestic and wild sheep and goats in Iran. *Small. Rum. Res.* 209, 106641. doi:10.1016/j.smallrumres.2022.106641
- Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS. Biol.* 5 (7), e171. doi:10.1371/journal.pbio.0050171
- Tao, L., He, X., Jiang, Y., Liu, Y., Ouyang, Y., Shen, Y., et al. (2021a). Genome-wide analyses reveal genetic convergence of prolificacy between goats and sheep. *Genes* 12 (4), 480. doi:10.3390/genes12040480
- Tao, L., He, X., Wang, F., Zhong, Y., Pan, L., Wang, X., et al. (2020). Luzhong mutton sheep: inbreeding and selection signatures. *J. Anim. Sci. Technol.* 62 (6), 777–789. doi:10.5187/jast.2020.62.777
- Tao, L., He, X., Wang, X., Di, R., and Chu, M. (2021b). Litter size of sheep (*Ovis aries*): inbreeding depression and homozygous regions. *Genes* 12 (1), 109. doi:10.3390/genes12010109
- Taylor, D. L., Zhong, L., Begg, D. J., de Silva, K., and Whittington, R. J. (2008). Toll-like receptor genes are differentially expressed at the sites of infection during the progression of John's disease in outbred sheep. *Vet. Immunol. Immunopathol.* 124 (1–2), 132–151. doi:10.1016/j.vetimm.2008.02.021
- Thameem, F., Yong, X., Permana, P. A., Wolford, J. K., Bogardus, C., and Prochazka, M. (2003). Evaluation of the microsomal glutathione S-transferase 3 (*MGST3*) locus on 1q23 as a Type 2 diabetes susceptibility gene in Pima Indians. *Hum. Genet.* 113 (4), 353–358. doi:10.1007/s00439-003-0980-y
- Tian, Y. Z., Usman, T., Tian, K. C., Di, J., Huang, X. X., and Xu, X. M. (2017). Comparative study of 13 candidate genes applying multi-reference normalization to detect the expression of different fineness in skin tissues of wool sheep. *Genet. Mol. Res.* 16 (10.4238).
- 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 (3), e72. doi:10.1371/journal.pbio.0040072
- Wang, B., Guo, H., Yu, H., Chen, Y., Xu, H., and Zhao, G. (2021). The role of the transcription factor EGR1 in cancer. *Front. Oncol.* 11, 642547. doi:10.3389/fonc.2021.642547
- Wang, Y., Zhang, C., Wang, N., Li, Z., Heller, R., Liu, R., et al. (2019). Genetic basis of ruminant headgear and rapid antler regeneration. *Science* 364 (6446), eaav6335. doi:10.1126/science.aav6335
- Xu, S. S., Gao, L., and Lyu, F. (2021). Whole-genome selective scans detect genes associated with important phenotypic traits in sheep (*Ovis aries*). *Front. Genet.* 12, 738879. doi:10.3389/fgene.2021.738879
- Xu, S. S., Gao, L., Ren, Y. L., Wang, F., Eypórsdóttir, E., Hallsson, J. H., et al. (2018). Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Front. Genet.* 9, 340573. doi:10.3389/fgene.2018.00118

- Yang, F., Wang, M., Zhang, B., Xiang, W., Zhang, K., Chu, M., et al. (2018). Identification of new progestogen-associated networks in mammalian ovulation using bioinformatics. *Bmc. Syst. Biol.* 12, 36–12. doi:10.1186/s12918-018-0577-7
- Zeder, M. A. (1999). Animal domestication in the Zagros: a review of past and current research. *Paléorient* 25, 11–25. doi:10.3406/paleo.1999.4684
- Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *Proc. Nation. Acad. Sci.* 105 (33), 11597–11604. doi:10.1073/pnas.0801317105
- Zeraatpisheh, Y., Zerehdaran, S., and Javadmanesh, A. (2023). Investigation of metabolic pathways related to the QTLs of parasite resistance trait in sheep genome using gene network and gene ontology. *Vet. Res. Biolo. Prod.* 36 (1), 83–90. In Persian. doi:10.22092/vj.2022.357660.1941
- Zhang, C. L., Liu, C., Zhang, J., Zheng, L., Chang, Q., Cui, Z., et al. (2022). Analysis on the desert adaptability of indigenous sheep in the southern edge of Taklimakan Desert. *Sci. Rep.* 12 (1), 12264. doi:10.1038/s41598-022-15986-x
- Zhang, Z., Tian, C., Zhou, S., Wang, W., Guo, Y., Xia, J., et al. (2012). Mechanism-based design, synthesis and biological studies of N⁹-substituted tetrahydrofolate analogs as inhibitors of cobalamin-dependent methionine synthase and potential anticancer agents. *Eur. J. Med. Chem.* 58, 228–236. doi:10.1016/j.ejmech.2012.09.027



OPEN ACCESS

EDITED BY

Fei Hao,
Northumbria University, United Kingdom

REVIEWED BY

Teng Xu,
Wenzhou University, China
Xin Wang,
Northwest A&F University, China

*CORRESPONDENCE

Dongjun Liu
✉ nmliudongjun@sina.com

RECEIVED 10 December 2024

ACCEPTED 20 January 2025

PUBLISHED 25 February 2025

CITATION

Han X, Gao G, Sun N, Dai B, Ren L, Bai H,
Liu J, Liu J, Zhao H and Liu D (2025)
Comparative proteomic analysis of the
telogen-to-anagen transition in cashmere
goat secondary hair follicles.
Front. Vet. Sci. 12:1542682.
doi: 10.3389/fvets.2025.1542682

COPYRIGHT

© 2025 Han, Gao, Sun, Dai, Ren, Bai, Liu, Liu,
Zhao 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.

Comparative proteomic analysis of the telogen-to-anagen transition in cashmere goat secondary hair follicles

Xiaoyu Han¹, Guizhen Gao^{1,2}, Nanxiang Sun¹, Bai Dai^{1,3},
Liqing Ren¹, Haobing Bai^{1,4}, Jianing Liu¹, Jie Liu^{1,5},
Hongyan Zhao⁶ and Dongjun Liu^{1*}

¹Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot, China, ²Peking University Cancer Hospital, Affiliated Cancer Hospital of Inner Mongolia Medical University, Hohhot, China, ³Reproductive Medicine Center, the Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China, ⁴Inner Mongolia People's Hospital NHC Key Laboratory of Diagnosis and Treatment of COPD, Inner Mongolia Key Laboratory of Respiratory Diseases, Hohhot, China, ⁵Tongliao Institute of Agriculture and Animal Husbandry, Tongliao, China, ⁶Agriculture and Animal Husbandry Technology Extension Center, Etuoke Banner, China

Secondary hair follicles (SHFs) in cashmere goats produce high-value cashmere fibers, which cyclic regulation is critical for optimizing cashmere yield and quality. This study explores the phenotypic changes and differential protein expression profiles involved in the telogen-to-anagen transition of SHFs. Through histological observations, proteomic analyses, and immunohistochemical validation, we identified key molecular features and regulatory pathways underlying SHF cyclic renewal. Histological analysis showed that telogen-phase SHFs exhibit a reduced volume, decreased dermal papilla cell (DPC) and hair matrix cell (HMC) activity, compact structure, and superficial localization in the dermis. Anagen-phase SHFs exhibit significantly increased volume, deeper dermal penetration, and active cell proliferation. Proteomic analysis identified 3,654 proteins in skin samples, with 458 differentially expressed proteins (DEPs) significantly associated with biological processes such as cell adhesion, signal transduction, protein synthesis, and metabolism. These DEPs were enriched in key regulatory pathways, including Notch, Wnt, Jak-STAT, PI3K-Akt, and ECM-receptor interaction. Protein-protein interaction analysis identified A Disintegrin and Metalloproteinase Domain 17 (ADAM17), Secreted Frizzled-Related Protein 1 (SFRP1), and Protein Phosphatase 1 Catalytic Subunit Alpha (PPP1CA) as core regulators of SHF cyclic transitions. Validation by RT-qPCR, Western blot, and immunohistochemical analyses confirmed that ADAM17, SFRP1, and PPP1CA were predominantly localized in functional regions, including the outer root sheath (ORS), dermal papilla (DP), and hair matrix (HM). Their expression levels were significantly enhanced during anagen. ADAM17 is suggested to promote the growth of SHFs by regulating ORS cells proliferation and mediating signal transduction in DPCs, while SFRP1, as a modulator of the Wnt signaling pathway, likely supports SHFs growth and regeneration by modulating the activity of Secondary hair follicle stem cells (SHFSCs) and promoting the differentiation of HMCs. PPP1CA may enhance cell proliferation and metabolic activity by modulating phosphorylation states. In conclusion, this study identifies key molecular factors and pathways driving the telogen-to-anagen transition in cashmere goat SHFs. It emphasizes the roles of ADAM17, SFRP1, and PPP1CA in SHF renewal and offers insights into SHF development mechanisms and cashmere fiber improvement.

KEYWORDS

cashmere goat secondary hair follicles, telogen-to-anagen transition, proteomic analysis, ADAM17, Sfrp1, PPP1CA, secondary hair follicle cycle regulation

1 Introduction

SHFs in cashmere goats produce cashmere, a highly valuable natural fiber renowned for its exceptional softness, warmth, and lightness. The growth and shedding of these fibers are tightly regulated by the cyclical activity of SHFs, alternating between telogen and anagen phases (1). Understanding the molecular mechanisms that govern the transition of the SHF cycle is essential for optimizing both cashmere yield and quality.

The hair follicle cycle comprises three distinct phases: anagen, catagen, and telogen. This finely tuned process is orchestrated by the complex interplay of multiple signaling pathways, which regulate the activation of hair follicle stem cells (HFSCs), the function of DPCs, and the dynamic remodeling of the extracellular matrix (ECM). These mechanisms collectively control the transitions between different phases of the follicular cycle (2–4). Among these pathways, the Wnt signaling pathway plays a central role by activating HFSCs to drive the transition from telogen to anagen, while its dysregulation can impair follicular regeneration (5–8). The Notch signaling pathway maintains the balance between HFSC self-renewal and differentiation (9), working synergistically with Wnt signaling to ensure proper follicular regeneration (10). Dysregulated Notch signaling can delay follicular cycle progression (11). The Ras-MAPK signaling pathway, particularly through its ERK and p38 branches, regulates the proliferation and differentiation of DPCs, making it critical for the anagen phase (12–14). Meanwhile, the ECM-receptor interaction pathway sustains follicular microenvironment homeostasis by providing structural support and transmitting biochemical signals, playing a vital role in cycle transitions (15). The Jak-STAT signaling pathway mediates cytokine-driven signal transduction to maintain the quiescent state of hair follicles, and its inhibition facilitates the induction of hair growth from dormant follicles (16, 17). Additionally, the Hippo pathway regulates the proliferation and differentiation of HFSCs through its downstream effectors YAP/TAZ (18). The PI3K-Akt pathway supports anagen by regulating cell survival, proliferation, and metabolism (19), while the TNF signaling pathway promotes the transition from telogen to anagen by activating HFSCs through the AKT/ β -catenin pathway (20). Despite advancements in understanding the roles of these signaling pathways in hair follicle cycling, the specific molecular regulators and their dynamic expression patterns during the telogen-to-anagen transition in the SHFs of cashmere goats remain poorly understood.

This study investigates the phenotypic and molecular changes underlying the telogen-to-anagen transition in cashmere goat SHFs. By integrating histological observations, proteomic analyses, and the validation and localization of key proteins, we comprehensively uncover the dynamic regulatory mechanisms governing the cycling of SHFs. Specifically, this study highlights the critical roles of ADAM17, SFRP1, and PPP1CA in regulating the telogen-to-anagen transition and promoting the growth of cashmere fibers. These findings enhance our understanding of hair follicle biology and provide a theoretical foundation for improving cashmere production through molecular breeding or biotechnological approaches.

2 Materials and methods

2.1 Experimental animals and sample collection

Six healthy adult cashmere goats ($n = 6$, aged 1 year; 3 females and 3 males) were selected from the Yiwei White Cashmere Goat Farm in Ordos, Inner Mongolia, China. All animals were maintained under standard feeding conditions with unrestricted access to food and water. Skin tissue samples were collected from the dorsal, flank, and abdominal regions during the telogen phase (April) and anagen phase (September) of the SHF cycle. These regions were chosen because they represent the primary distribution areas of SHFs and exhibit significant differences in density, developmental state, and cycle activity. The dorsal region has the highest SHF density and the most active anagen-phase follicles, making it the primary source of high-quality cashmere fibers. The flank region shows intermediate follicle density and activity, with structural characteristics between those of the dorsal and abdominal regions. The abdominal region has the lowest follicle density, and some SHFs may be in a degenerated state. Sampling these regions provides a comprehensive representation of the periodic changes in SHFs across the goat's body and offers a basis for exploring region-specific regulatory mechanisms. The phases of the SHF cycle were determined through histological examination and follicular morphology assessment. The collected skin samples ($\sim 1 \text{ cm}^2$) were immediately divided into two parts: one part was fixed in 4% paraformaldehyde for histological and immunohistochemical analyses, while the other part was rapidly frozen in liquid nitrogen for RNA and protein extraction, proteomic sequencing, and subsequent analyses.

2.2 Experimental design

As shown in Figure 1, this study employed tandem mass tag (TMT)-based labeling technology combined with LC-MS/MS to identify and quantify the proteome of cashmere goat SHFs during the telogen and anagen phases. Skin tissue samples from six goats were used: three females (TelF1, TelF2, TelF3) and three males (TelM1, TelM2, TelM3) during telogen, and the same three females (AnaF1, AnaF2, AnaF3) and three males (AnaM1, AnaM2, AnaM3) during anagen. Samples were divided into two independent experimental groups for analysis. Proteins were first extracted from skin at both telogen and anagen phases. Samples underwent reduction and alkylation of cysteine residues to minimize disulfide bond interference. Proteins were then digested into peptides using trypsin. The peptide samples were labeled with TMT reagents (TMT10plex), with TMT tags assigned to telogen and anagen samples in each experiment, alongside a reference sample (Ref) composed of an equal mix of all experimental samples. The labeled samples were mixed at a 1:1 ratio for subsequent analysis. The mixed samples were analyzed using liquid chromatography–tandem mass spectrometry (LC-MS/MS) for protein identification and

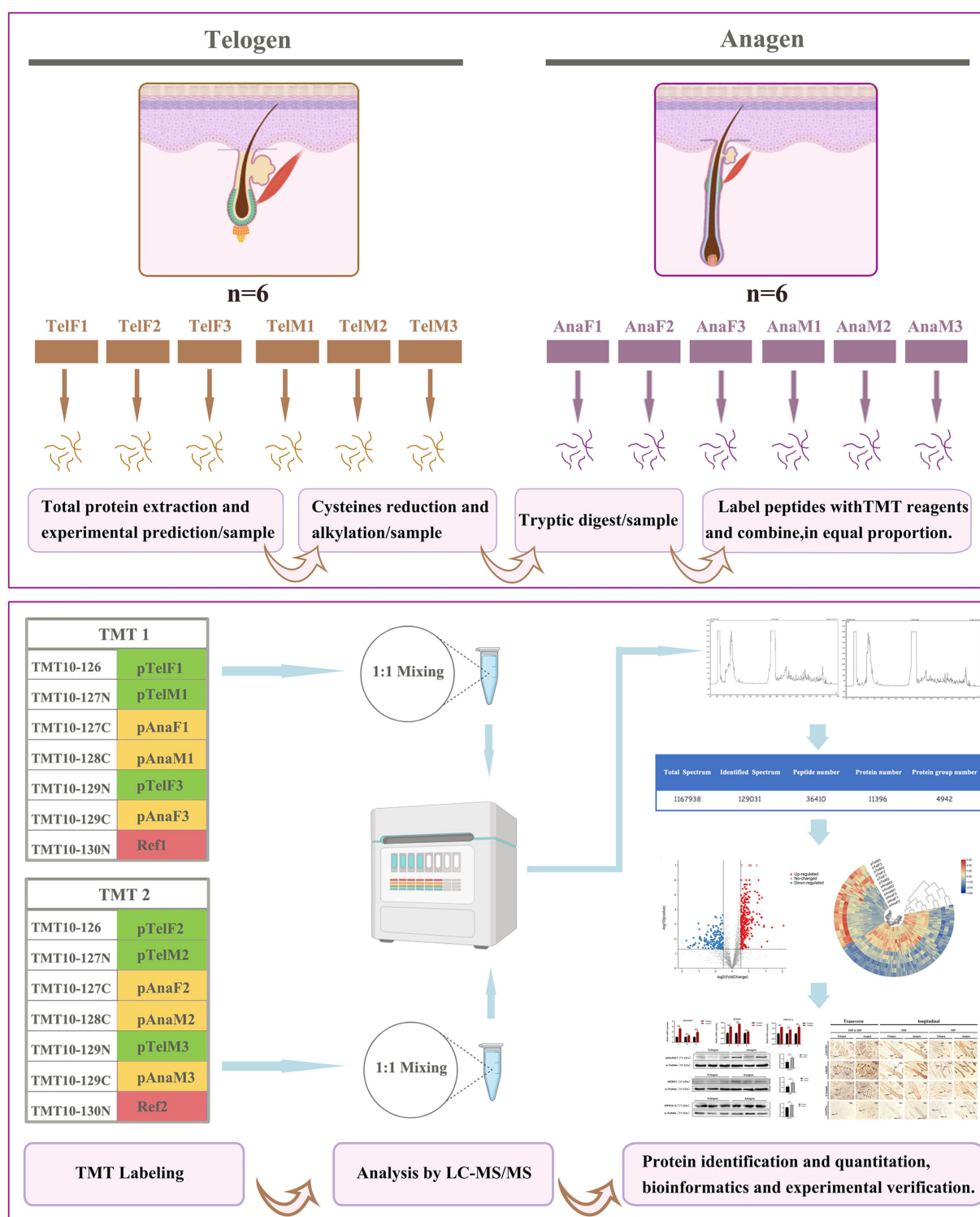


FIGURE 1

Workflow of TMT-labeled proteomics analysis of SHFs during the telogen- and anagen-phase transitions. This study utilized TMT labeling combined with LC-MS/MS to identify and quantify the proteome of cashmere goat SHFs during the telogen and anagen phases. Skin tissue samples included six telogen samples (three females and three males) and six anagen samples (the same three females and three males). After protein extraction, cysteine reduction, and alkylation treatment, the proteins were digested into peptides and labeled with TMT reagents. The labeled samples were mixed in a 1:1 ratio and analyzed via LC-MS/MS for protein identification and quantification. DEPs were further analyzed using bioinformatics to elucidate their functions, while the expression patterns and localization of key proteins were experimentally validated.

quantification. Mass spectrometry (MS) data were used to generate total spectra, identified spectra, peptide counts, and the protein

dataset. Proteomic data were subjected to statistical analysis to identify DEPs, which were then functionally annotated and

enriched. Correlation analyses were conducted to further explore biological differences between telogen and anagen. Additionally, key DEPs were validated using real-time quantitative PCR (RT-qPCR) and Western blotting, and their expression patterns and localization were evaluated via immunohistochemistry.

2.3 Histological and immunohistochemical analysis

Skin samples were fixed in 4% paraformaldehyde, dehydrated, embedded in paraffin, and sectioned at a thickness of 7 μm . The sections were stained with hematoxylin and eosin (H&E) to observe SHF structure and confirm the telogen or anagen phase. For immunohistochemical analysis, paraffin sections were deparaffinized, rehydrated, and subjected to antigen retrieval in boiling sodium citrate buffer. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide. Sections were incubated overnight at 4°C with primary antibodies. The following day, the sections were incubated with HRP-conjugated secondary antibodies for 1 h, followed by DAB staining. Images were captured using an optical microscope (Nikon, Tokyo, Japan). Detailed antibody information is provided in [Supplementary Table S1](#).

2.4 RNA extraction and RT-qPCR

Total RNA was extracted from skin samples using RNAiso reagent (Takara Bio Inc., Shiga, Japan) following the manufacturer's protocol. RNA purity and concentration were measured using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). Complementary DNA (cDNA) was synthesized using the PrimeScript FAST reverse transcription (RT) reagent kit with genomic DNA (gDNA) eraser (Takara Bio Inc., Shiga, Japan). Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) was performed on a CFX96 Real-Time PCR System (Bio-Rad Laboratories, Hercules, CA, United States) with TB Green® Premix Ex Taq™ II (Takara Bio Inc., Shiga, Japan). Specific primers were designed for ADAM17, SFRP1, and PPP1CA, with GAPDH as the reference gene for relative expression calculation using the $2^{-\Delta\Delta\text{Ct}}$ method. Primer sequences are listed in [Supplementary Table S2](#).

2.5 Protein extraction and western blot

Total protein was extracted from skin samples using a Mammalian Protein Extraction Kit (CWBIO, Beijing, China) according to the manufacturer's instructions. Protein concentration was determined using a BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, WA, United States). Equal amounts of protein (20 μg per lane) were separated by SDS-PAGE and transferred to PVDF membranes. Membranes were blocked with 5% non-fat milk at room temperature for 1 h and incubated overnight at 4°C with primary antibodies. After incubation with HRP-conjugated secondary antibodies for 1 h, protein bands were visualized using the Tanon 5,200 Imaging System (Tanon, Shanghai, China). Band intensity was quantified using ImageJ software, and statistical analysis was performed to compare protein

expression levels across groups. Detailed antibody information is provided in [Supplementary Table S3](#).

2.6 Proteomic sequencing and bioinformatic analysis

Following protein extraction, protein solutions were reduced with 10 mmol/L TCEP at 37°C for 1 h and alkylated with 40 mmol/L iodoacetamide at room temperature in the dark for 30 min. Trypsin (Promega, Madison, WI, United States) was added at an enzyme-to-protein ratio of 1:50 (w/w) for overnight digestion at 37°C. The resulting peptides were labeled with TMT reagents (Thermo Fisher Scientific, Waltham, MA, United States) and incubated at room temperature for 2 h, followed by a 15-min reaction with hydroxylamine. Equal amounts of labeled peptides were mixed, vacuum-dried, re-dissolved in 0.1% formic acid, and analyzed using a Thermo Scientific Q Exactive mass spectrometer coupled with an EASY-nLC 1,200 nano-liquid chromatography system.

Peptide samples were loaded onto a C18 column (75 $\mu\text{m} \times 25\text{ cm}$, Thermo Fisher Scientific, Waltham, MA, United States) and separated using gradient elution with acetonitrile (ACN) containing 0.1% formic acid at a flow rate of 300 $\mu\text{L}/\text{min}$. The mass spectrometer was operated in data-dependent acquisition (DDA) mode, with a full scan range of 350–1,300 m/z at resolutions of 70,000 (full scan) and 35,000 (MS/MS scan), followed by MS/MS analysis of the top 20 most abundant precursor ions. Raw mass spectrometry data were analyzed using Proteome Discoverer™ Software 2.2 against the UniProt goat protein database. The false discovery rate (FDR) for peptide and protein identification was set to $\leq 1\%$, with proteins identified by at least one unique peptide.

TMT labeling technology, with the quantification based on the intensity of the reporter ions. To correct for potential technical variances, normalization of the protein expression levels was conducted across samples using total ion intensity normalization. Additionally, inter-TMT batch normalization was performed using an internal reference sample, ensuring consistent protein quantification across independent TMT experiments. This step minimized batch effects and allowed for accurate comparison of protein expression levels across all samples.

Differential protein expression was analyzed using the *R* statistical package with the *t*-test function. DEPs were defined as upregulated [$p < 0.05$ and fold change (FC) > 1.2] or downregulated ($p < 0.05$ and $\text{FC} < 0.83$). Heatmaps of DEPs were generated using the *R* package pheatmap. Functional annotation of DEPs was performed using Gene Ontology (GO), while Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was conducted using the clusterProfiler *R* package. Protein–protein interaction (PPI) networks were constructed using the STRING database and visualized with Cytoscape software.

2.7 Statistical analysis

All data are presented as means $\pm$ standard deviation (SD). Student's *t*-test was used for significance testing between two groups, while one-way analysis of variance (ANOVA) was applied for comparisons among three or more groups. A significance threshold of $p < 0.05$ was used.

3 Results

3.1 Significant phenotypic changes observed during the telogen-to-anagen transition of SHFs

Skin samples from the dorsal, flank, and abdominal regions of cashmere goats were examined using hematoxylin and eosin (H&E) staining, along with transverse and longitudinal observations. These analyses revealed significant phenotypic changes in SHFs during the telogen-to-anagen transition, as well as notable structural differences among the regions.

During the telogen phase (Figure 2A), SHFs exhibited significant phenotypic changes, including a substantial reduction in follicle volume, a decrease in number, more superficial positioning within the skin, and a more compact structure. Moreover, the activity of DPCs and HMCs was markedly reduced. Transverse sections revealed that dorsal SHFs had the highest density, compact structures, and relatively well-preserved morphology; flank SHFs showed moderate density and regular morphology; while abdominal SHFs had the lowest density, larger cross-sectional areas, and more pronounced degeneration of secondary structures. Longitudinal sections further demonstrated that dorsal SHFs, despite being positioned closer to the epidermis, maintained greater depth and structural integrity. In contrast,

abdominal SHFs were more superficial, with severe degeneration of secondary structures and nearly complete cessation of hair growth activity.

During the anagen phase (Figure 2B), SHFs exhibited significant phenotypic changes. The follicle volume increased substantially, structures became more distinct, and follicles extended deeper into the skin. Dorsal SHFs showed the most pronounced growth, with follicle bases reaching the deep dermis, the highest proliferative activity in HMCs, and well-developed SHFs contributing to robust hair growth. Flank SHFs displayed intermediate growth, with follicle bases extending to the mid-dermis and moderately developed structures. Abdominal SHFs exhibited the smallest increase in volume and depth, extending only to the superficial dermis. Poorly developed SHFs in this region resulted in reduced hair production. Transverse sections revealed a significant increase in the number and diameter of SHFs in the dorsal region, followed by the flank, while abdominal SHFs showed the smallest increase. Longitudinal sections further demonstrated that dorsal hair papillae and ORS structures were the most well-developed, whereas abdominal SHFs displayed relatively simple and underdeveloped structures.

In summary, the phenotypic differences between telogen and anagen SHFs were reflected in their size, position, developmental state, and cellular activity. Furthermore, SHFs in the dorsal, flank, and abdominal regions exhibited distinct structural and developmental characteristics. These region-specific features provide an important

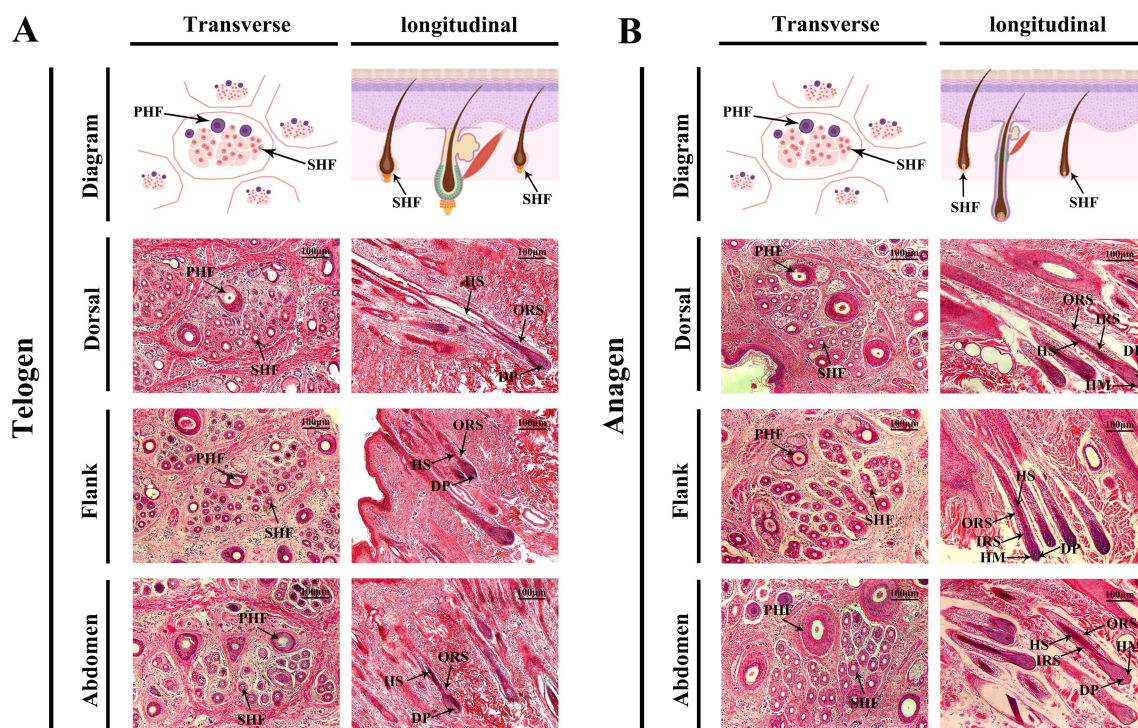


FIGURE 2

Morphological changes in cashmere goat SHFs during the telogen and anagen phases. (A) Representative cross-sections and longitudinal sections of SHFs in the telogen phase. Transverse sections show primary and secondary hair follicles in the dorsal, flank, and abdominal regions. Telogen-phase SHFs appear significantly shrunken, with compact, round, or oval structures. Longitudinal sections display shorter telogen-phase SHFs with their bases located close to the epidermis. The DP and HM are underdeveloped, and the overall follicular structure is degenerated. (B) Representative cross-sections and longitudinal sections of SHFs in the anagen phase. Transverse sections reveal enlarged SHFs in the dorsal, flank, and abdominal regions, with increased diameter and well-defined structures. Longitudinal sections demonstrate the deep extension of SHFs into the dermis, with well-developed DP, HM, IRS, and ORS. Scale bar: 100 μm.

foundation for further investigations into the molecular mechanisms underlying the telogen-to-anagen transition and the functional diversity of SHFs across body regions.

3.2 Proteomic analysis reveals characteristic protein expression during the telogen-to-anagen transition of SHFs

To elucidate protein expression profiles during the transition of SHFs from the telogen-to-anagen phase, high-throughput proteomic analysis was performed on skin samples from cashmere goat. Telogen-phase skin samples were used as the control group, while anagen-phase samples constituted the experimental group. Each group included six biological replicates (three male and three female goats), with all samples collected from the same six goats to ensure consistency.

Following protein extraction, enzymatic digestion, liquid chromatography-mass spectrometry (LC-MS) analysis, and quality control, a total of 3,654 proteins were identified through database matching (Supplementary Table S4). These identified proteins were associated with multiple key processes related to hair follicle growth and metabolism, providing a foundation for subsequent analysis of DEPs and their roles in regulating the cycle of SHFs.

3.2.1 Data correlation analysis

To evaluate the effects of different growth stages and sexes on protein expression in SHFs, principal component analysis (PCA) was performed on proteomic data from telogen-phase females (pTelF), anagen-phase females (pAnaF), telogen-phase males (pTelM), and anagen-phase males (pAnaM).

For female samples, PCA revealed that PC1 and PC2 accounted for 37.40 and 22.20% of the total variation, respectively.

Telogen-phase samples (pTelF) and anagen-phase samples (pAnaF) were clearly separated along the PC1 axis, with anagen-phase samples clustering tightly. This clustering indicated high within-group consistency in protein expression during anagen. In contrast, telogen-phase samples were more dispersed, reflecting greater heterogeneity in protein expression during this phase (Figure 3A).

For male samples, PCA showed that PC1 and PC2 explained 40.50 and 22.40% of the total variation, respectively. Similarly, telogen-phase samples (pTelM) and anagen-phase samples (pAnaM) were distinctly separated along the PC1 axis. Anagen-phase male samples also clustered tightly, further supporting the observation of consistent protein expression during anagen, while telogen-phase samples exhibited greater dispersion (Figure 3B).

Overall, PCA results indicated significant differences in protein expression between telogen and anagen phases, regardless of sex. Protein expression in anagen-phase samples was more consistent, while telogen-phase samples displayed greater heterogeneity. These findings provide a foundation for further exploration of DEPs and their regulatory mechanisms in the SHF growth cycle.

3.2.2 Functional annotation and enrichment analysis of DEPs

To identify changes in protein expression during the transition of SHFs from the telogen-to-anagen phase, a two-tailed Student's *t*-test was performed to evaluate statistical significance, with a $p < 0.05$ considered significant. Proteins with a fold change (FC) > 1.2 were classified as upregulated, while those with an FC < 0.83 were classified as downregulated. A total of 458 DEPs were identified (Supplementary Table S5), including 293 upregulated and 165 downregulated proteins, indicating significant differences in protein expression between the two phases of SHFs.

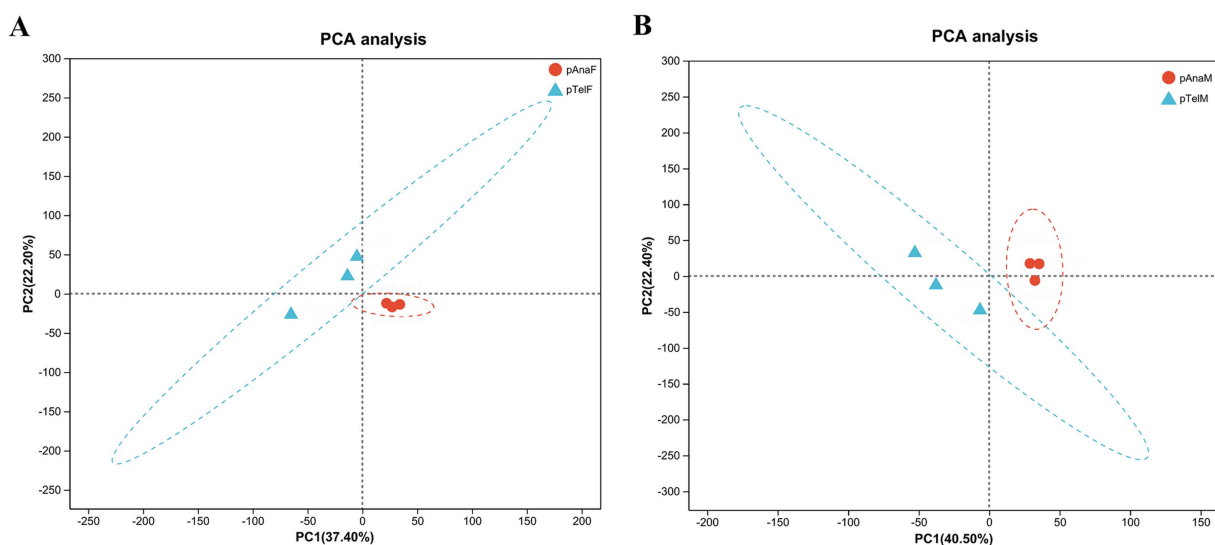


FIGURE 3

PCA of protein expression profiles in telogen- and anagen-phase SHFs of cashmere goats. **(A)** PCA of protein expression in telogen-phase (pTelF) and anagen-phase (pAnaF) female goats. PC1 (37.40%) and PC2 (22.20%) account for the majority of the variation. Telogen-phase samples (blue triangles) are distinctly separated from anagen-phase samples (red circles) along the PC1 axis. Anagen-phase samples exhibit tighter clustering, suggesting greater consistency in protein expression. **(B)** PCA of protein expression in telogen-phase (pTelM) and anagen-phase (pAnaM) male goats. PC1 (40.50%) and PC2 (22.40%) account for the majority of the variation. Telogen-phase samples (blue triangles) are clearly separated from anagen-phase samples (red circles) along the PC1 axis, with anagen-phase samples showing tighter clustering, indicating higher consistency in protein expression.

A volcano plot displayed the distribution of DEPs, illustrating that numerous proteins were significantly upregulated in anagen-phase SHFs, while some were markedly downregulated (Figure 4A). A heatmap of the DEPs revealed distinct clustering patterns, indicating clear grouping trends in protein expression between the two phases of SHFs (Figure 4B).

To investigate the biological functions and signaling pathways involved in the transition from the telogen-to-anagen phase of SHFs, GO functional annotation and KEGG pathway enrichment analyses were conducted on the identified DEPs. GO analysis classified DEPs into three main categories: biological processes (BP), cellular components (CC), and molecular functions (MF).

In the BP category, DEPs were significantly enriched in processes such as rRNA metabolism, RNA catabolic processes, and protein targeting to the endoplasmic reticulum, indicating their critical roles in regulating cellular transcription and translation. In the CC category, DEPs were primarily enriched in ribosomal subunits (both large and small) and intermediate filaments, suggesting their association with cellular organelles and cytoskeletal structures essential for SHF function and morphology. In the MF category, DEPs were significantly enriched in RNA binding, structural molecule activity, and oxygen transport activity, highlighting their involvement in cellular metabolism, protein synthesis, and oxygen delivery (Figure 5A).

KEGG pathway enrichment analysis revealed several pathways significantly associated with the transition of SHFs from the telogen-to-anagen phase. Notably, the ribosome pathway, directly associated with protein translation and synthesis, suggested heightened metabolic activity in anagen-phase SHFs. Additionally, the purine metabolism and riboflavin metabolism pathways reflected the increased demand for energy and nucleic acid metabolism to support SHF growth. The PPAR signaling and

Notch signaling pathways, both well-known regulators of hair follicle cycling and cell differentiation, were also significantly enriched. Other enriched pathways, including complement and coagulation cascades, myocardial contraction, and tyrosine metabolism, may reflect the metabolic regulation and growth environment required for SHF development and regeneration (Figure 5B).

To further explore the key pathways and protein functions involved in the regulation of the SHF cycle in cashmere goats, we selected pathways related to hair follicle development from the KEGG enrichment analysis of the 458 DEPs and identified a subset of 28 key proteins (Supplementary Table S6). GO functional annotation and KEGG pathway enrichment analyses were subsequently performed on this protein set.

GO analysis showed that in the BP category, DEPs were significantly enriched in processes such as the positive regulation of signal transduction, Notch receptor processing, cell adhesion, and biological adhesion. These findings suggest that these proteins may promote the transition of telogen SHFs to anagen by regulating intercellular signaling and interactions. In the CC category, DEPs were mainly localized in intercellular junctions, protein kinase CK2 complexes, and γ -secretase complexes, implying that these proteins play an important role in intercellular signal exchange and molecular complex regulation. In the MF category, DEPs were significantly enriched in macromolecular complex binding and protein phosphatase activity, indicating their critical roles in protein modification and signal pathway regulation (Figure 5C).

KEGG analysis showed that DEPs were significantly enriched in several signaling pathways closely related to hair follicle development, growth, and regeneration. These included the Notch signaling pathway, Wnt signaling pathway, and Hippo signaling pathway, all of

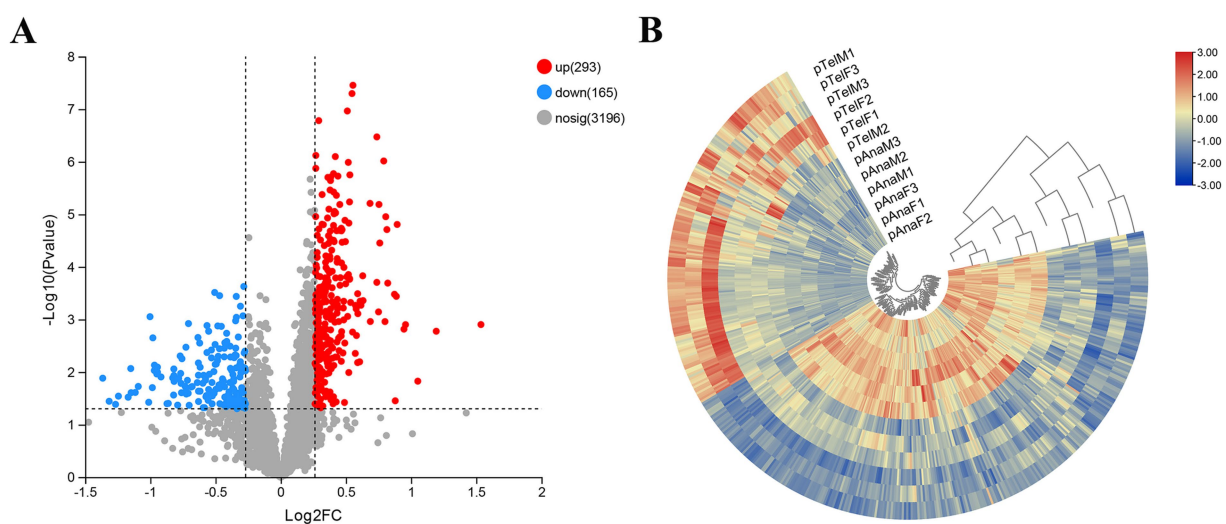
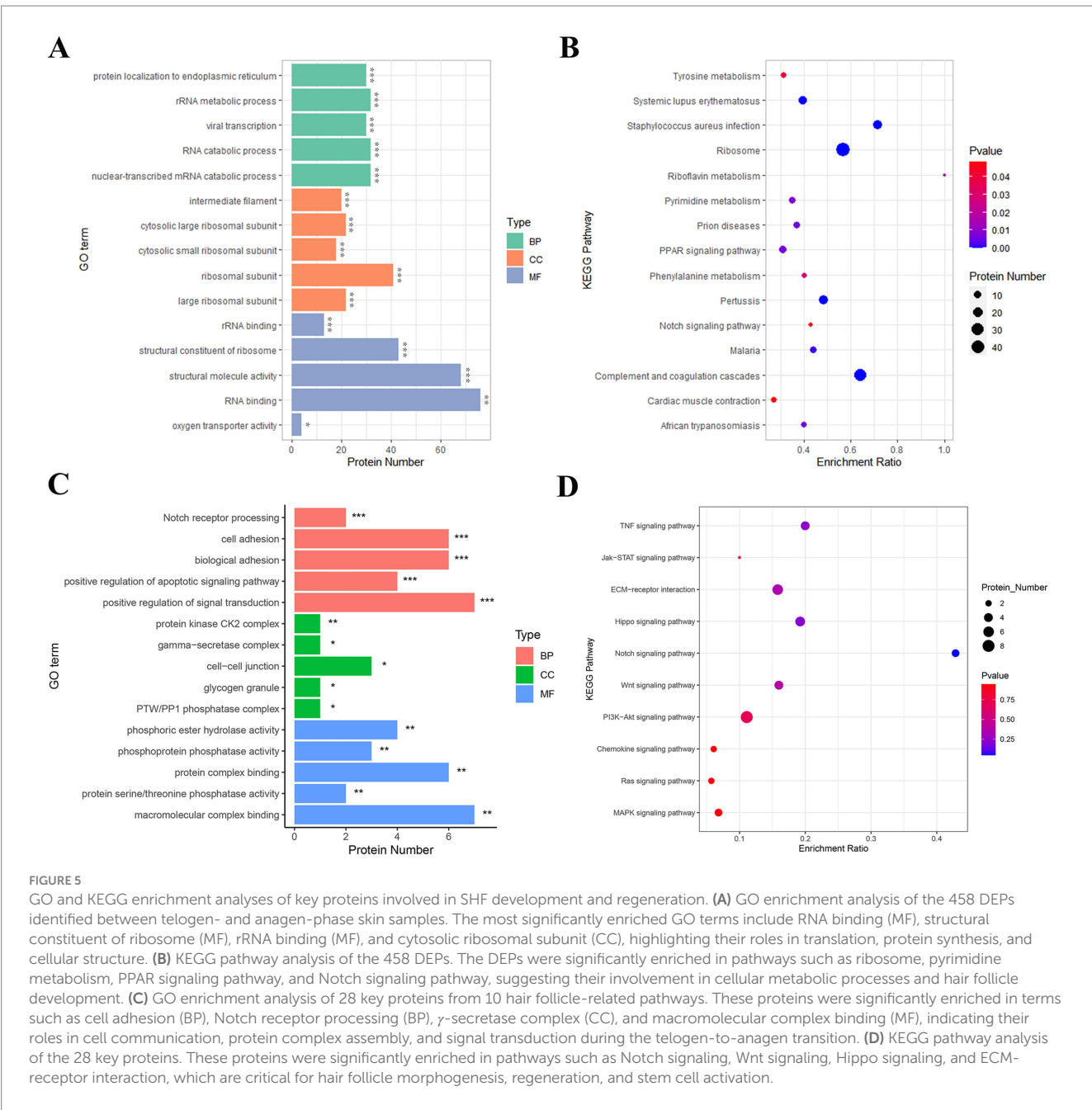


FIGURE 4

Identification and clustering of DEPs in telogen- and anagen-phase SHFs of cashmere goats. (A) Volcano plot illustrating the distribution of DEPs. The X-axis represents the log2(fold change) of protein expression, while the Y-axis represents the $-\log_{10}(p\text{-value})$. Proteins significantly upregulated ($\log_2\text{FC} > 0.32$, $p < 0.05$) are shown in red, whereas those significantly downregulated ($\log_2\text{FC} < -0.26$, $p < 0.05$) are shown in blue. Gray dots indicate proteins without significant expression changes. A total of 458 DEPs were identified, including 293 upregulated and 165 downregulated proteins. (B) Hierarchical clustering heatmap of DEPs. Samples include telogen-phase male goats (pTelM), telogen-phase female goats (pTelF), anagen-phase male goats (pAnaM), and anagen-phase female goats (pAnaF). Rows represent DEPs, while columns represent biological replicates. Red indicates upregulated proteins, blue indicates downregulated proteins, and yellow indicates proteins with no significant changes. Samples are distinctly clustered into telogen- and anagen-phase groups, highlighting differential protein expression patterns between the two phases.



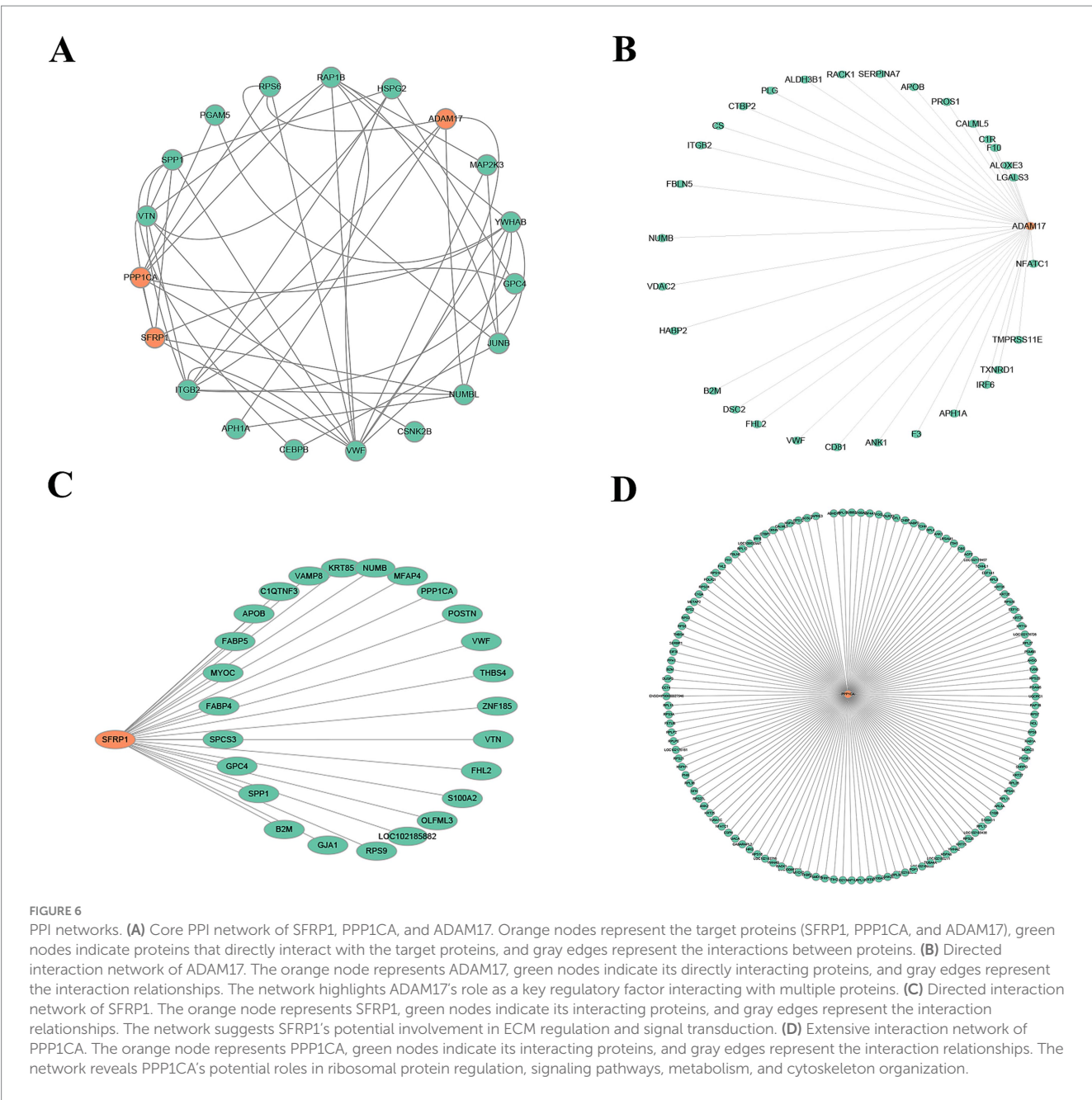
which are known to play central roles in hair follicle cycle regulation and stem cell activation. The PI3K-Akt signaling pathway and MAPK signaling pathway suggested that these proteins may be involved in initiating anagen by regulating cell proliferation and differentiation. ECM-receptor interaction and cytokine signaling pathways (such as TNF and Jak-STAT) further highlighted the importance of ECM remodeling and intercellular communication in the SHF microenvironment during the transition (Figure 5D).

In conclusion, an in-depth analysis of GO and KEGG indicates that the key DEPs involved in the telogen-to-anagen transition of SHFs are primarily associated with biological processes such as the positive regulation of signal transduction, cell adhesion, and protein modification. These proteins are significantly enriched in critical signaling pathways that regulate the hair follicle cycle. Collectively,

these findings elucidate the functions and molecular mechanisms of key DEPs during the telogen-to-anagen transition in cashmere goat SHFs, providing valuable theoretical evidence for future research into SHF cycle regulation.

3.2.3 Protein–protein interaction network analysis

To identify key factors involved in the periodic regulation of DEPs, a PPI network was constructed (Figure 6A). This network consisted of 19 nodes and multiple edges, representing the interactive relationships between proteins. Through topological analysis and functional annotation, three high-degree hub proteins—ADAM17, SFRP1, and PPP1CA—were identified as key regulators of the telogen-to-anagen transition in SHFs.



The interaction network of ADAM17 (Figure 6B) revealed its primary associations with ECM-related proteins, signal transduction-related proteins, metabolism- and oxidative stress-related proteins, and blood-related factors. This suggests that ADAM17 plays a critical role in remodeling the microenvironment of SHFs and mediating signaling pathways.

SFRP1 (Figure 6C) demonstrated close interactions with ECM-related proteins, metabolism-related proteins, signal transduction-related proteins, and cytoskeleton- and structure-related proteins. These findings indicate that SFRP1 is involved in regulating ECM signaling, maintaining structural integrity, and mediating metabolic processes.

The network of PPP1CA (Figure 6D) showed extensive interactions with ribosomal proteins, cytoskeleton-related proteins, signal regulation proteins, and metabolic proteins. This highlights

PPP1CA's role in modulating protein synthesis, cytoskeletal dynamics, and cellular signaling pathways.

Collectively, these findings suggest that ADAM17, SFRP1, and PPP1CA function as critical regulators in the telogen-to-anagen transition by influencing signal transduction, ECM remodeling, and metabolic processes. These results provide valuable insights into the molecular mechanisms underlying SHF cycle regulation.

3.3 Validation of ADAM17, SFRP1, and PPP1CA as key regulators during the telogen-to-anagen transition of SHFs

RT-qPCR and Western blot analyses confirmed that the expression levels of ADAM17, SFRP1, and PPP1CA were upregulated during the

telogen-to-anagen transition of SHFs (Figures 7A,B), consistent with the proteomic sequencing data. These findings suggest that these proteins may mediate the SHF cycle transition by regulating signal transduction and cell adhesion, providing a robust experimental foundation for further studies.

3.4 Immunolocalization of ADAM17, SFRP1, and PPP1CA in telogen and anagen phases of SHFs

Immunohistochemical staining was performed to analyze the expression and localization of ADAM17, SFRP1, and PPP1CA in SHFs from the dorsal region of cashmere goats during the telogen and anagen phases (Figure 8).

The immunolocalization and functional analysis of ADAM17 in transverse sections revealed its expression in SHFs during the telogen phase. During the anagen phase, its expression significantly increased in SHFs, particularly in the ORS. In longitudinal sections, ADAM17 exhibited weak expression in the ORS and DP of SHFs during telogen. However, its expression was markedly enhanced during anagen and localized primarily to the ORS and DP. These findings suggest that ADAM17 may facilitate the telogen-to-anagen transition by promoting ORS cells proliferation and mediating signal transduction in DPCs.

The immunolocalization and functional analysis of SFRP1 in transverse sections demonstrated its expression in SHFs during the telogen phase, which significantly increased during the anagen phase. Longitudinal sections revealed weak expression of SFRP1 in the ORS and HM of SHFs during the telogen phase. However, during the anagen phase, its expression was markedly enhanced and primarily

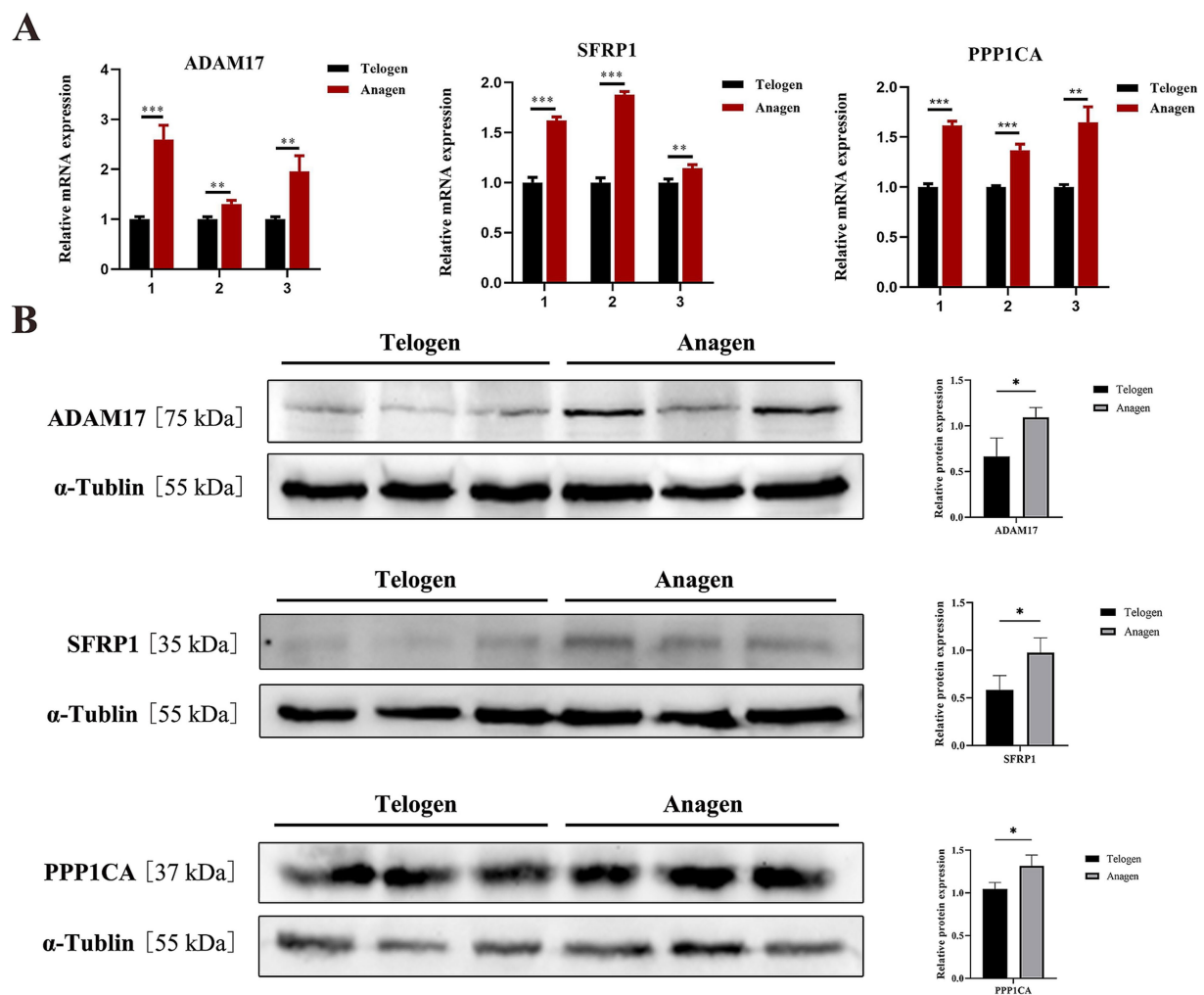


FIGURE 7

Validation of ADAM17, SFRP1, and PPP1CA as key regulators in the transition from telogen to anagen in SHFs. (A) Relative mRNA expression levels of ADAM17, SFRP1, and PPP1CA in telogen- and anagen-phase SHF samples, measured by RT-qPCR. The results show that the mRNA expression levels of all three genes are significantly higher in anagen samples compared to telogen samples. Samples 1, 2, and 3 represent skin tissues collected from the same three individual goats during the telogen and anagen phases (*** $p < 0.001$; ** $p < 0.01$). (B) Protein expression levels of ADAM17, SFRP1, and PPP1CA in telogen- and anagen-phase SHF samples, analyzed by Western blot with α -Tubulin as the internal control. Semi-quantitative analysis reveals that the protein expression levels of these three molecules are significantly higher in anagen samples than in telogen samples. Error bars represent the mean $\pm$ standard deviation (SD), with significance levels indicated as * $p < 0.05$.

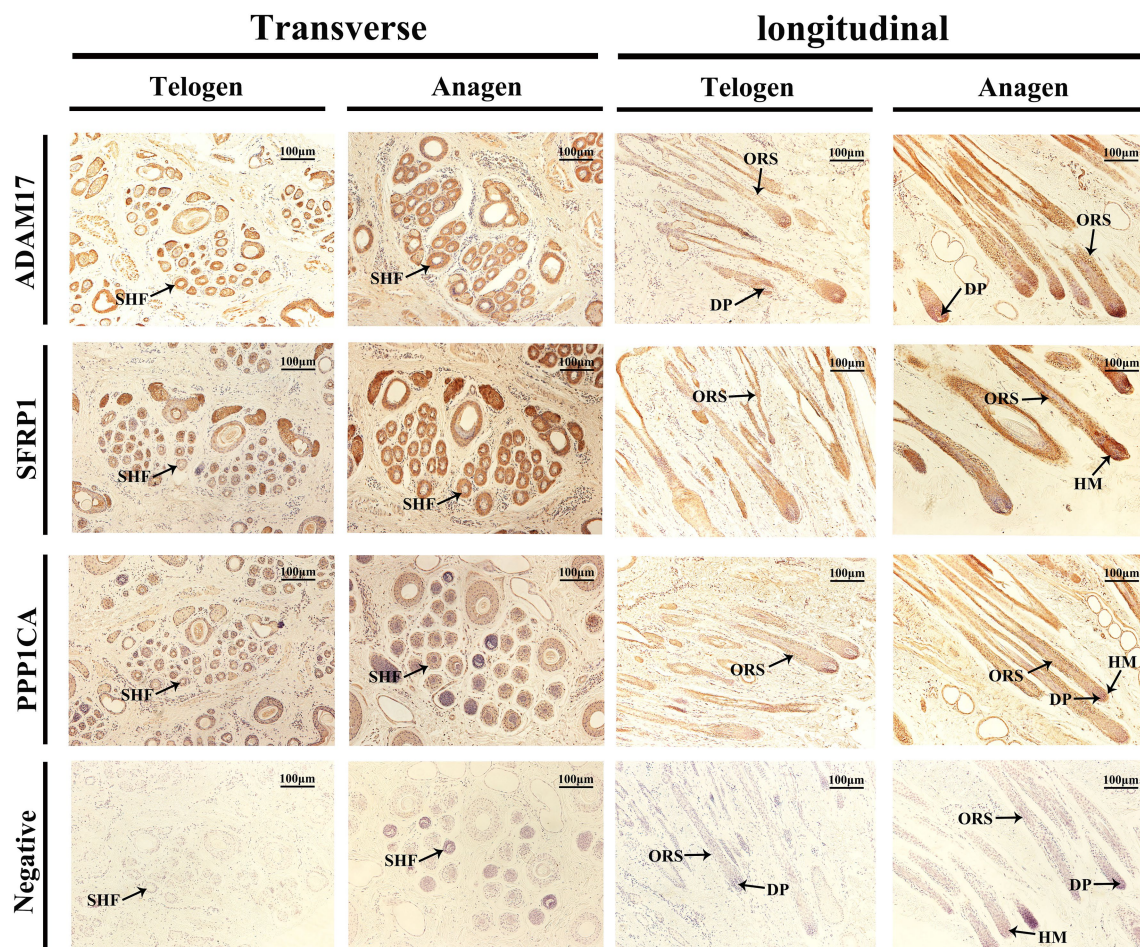


FIGURE 8

Immunolocalization of ADAM17, SFRP1, and PPP1CA in SHFs during the telogen and anagen phases. Immunohistochemical staining shows the expression and localization of ADAM17, SFRP1, and PPP1CA in transverse and longitudinal sections of SHFs from the dorsal region of cashmere goats during the telogen and anagen phases. Negative controls exhibit no staining, confirming the specificity of the antibodies. Scale bar: 100 μ m.

localized to the ORS and HM of SHFs. As an antagonist of the Wnt signaling pathway, SFRP1 likely supports SHF growth and regeneration by modulating the activity of SHFSCs and promoting the differentiation of HMCs.

The immunolocalization and functional analysis of PPP1CA in transverse sections revealed its expression in SHFs, with significantly heightened levels during the anagen phase, predominantly localized to the ORS. Longitudinal sections demonstrated weak expression of PPP1CA in the ORS during the telogen phase. However, during the anagen phase, its expression markedly increased and expanded to the ORS, HM, and DP. As a phosphatase, PPP1CA may promote the proliferation and differentiation of ORS cells, HMCs, and DPCs in SHFs by regulating the phosphorylation of cell cycle regulators and signaling components.

In summary, immunohistochemical analysis demonstrated dynamic and differential expression of ADAM17, SFRP1, and PPP1CA in hair follicles during the telogen and anagen phases, with primary localization to key regions such as the ORS, DP, and HM. These proteins likely regulate the SHF cycle by modulating signal transduction, stem cell activity, and cell proliferation. These findings provide important evidence for further exploration of the molecular mechanisms underlying SHF regulation.

4 Discussion

Hair follicles are important miniature organs in the skin, composed of both epidermal and dermal components. Their primary function is to produce hair through periodic activity, playing a vital role in thermoregulation, skin barrier formation, and environmental sensing in animals (21). The hair follicle cycle is divided into the anagen, catagen, and telogen. This dynamic periodicity is essential for hair growth, shedding, and regeneration (22, 23). The normal maintenance of the hair follicle cycle relies on a complex molecular signaling network, which is not only significant for understanding the regulatory mechanisms of the SHF cycle but also provides a theoretical basis for improving cashmere quality in cashmere goats.

Proteins, as pivotal molecules in cell signaling and functional regulation, play a central role in orchestrating transitions between different phases of the SHF cycle. Previous studies have primarily focused on the regulation of the hair follicle cycle by pathways such as Wnt, Notch, and Jak-STAT (2, 8, 16, 24). However, systematic studies on the functions, dynamic expression patterns, and mechanisms of specific proteins during different phases of the SHF cycle remain limited. While this study is constrained by a small sample size, we minimized individual variability by collecting telogen and anagen

skin samples from the same cashmere goat. Using proteomics, we successfully identified key proteins associated with the SHF cycle. This high-throughput proteomic approach provides a rapid, efficient, and comprehensive method to systematically analyze the dynamic changes in proteins during the SHF cycle, offering new insights into the molecular mechanisms underlying SHF regulation (25).

In this study, we performed LC-MS/MS-based proteomic analysis on skin tissues from telogen and anagen phases in cashmere goats, revealing dynamic protein changes during the telogen-to-anagen transition of SHFs. A large number of proteins were found to be either upregulated or downregulated, indicating significant protein remodeling during this phase, which is likely associated with the reconstruction of the non-permanent region of SHFs. Reconstruction of the non-permanent region is a critical turning point in hair follicle development and marks the initiation of the telogen-to-anagen transition (26). This event involves a series of complex biological processes, including cell proliferation and differentiation, ECM remodeling, signaling pathway regulation, and the activation of HFSCs (27–30). During this process, HFSCs must first transition from a quiescent to an active state, followed by amplification and differentiation to form new hair follicle structures (28). Concurrently, ECM remodeling provides mechanical support and signaling cues for hair follicle growth. Growth factors and cytokines play key roles in regulating HFSCs activation through signaling pathways (31–33). These mechanisms ensure the structural integrity and functional restoration of hair follicles. Additionally, HFSCs activation is tightly regulated by DP signals. As the “regulatory center” of the hair follicle cycle, the DP communicates bidirectionally with HFSCs by secreting signaling molecules, thereby promoting hair follicle regeneration (15).

Our findings revealed that ADAM17 was significantly upregulated during the anagen phase, suggesting that it may facilitate the activation of the telogen-to-anagen transition by cleaving membrane-bound growth factors. ADAM17 is a “shedase” that cleaves membrane-bound proteins to release their active forms, such as epidermal growth factor (EGF) and tumor necrosis factor- α (TNF- α) (34–37). Evidence suggests that EGFR ligands, such as EGF, promote the proliferation and migration of ORS cells (38–41). Additionally, the EGFR ligand EREG activates ORS cells and HMCs through EGFR signaling (42). These growth factors are critical for SHFSCs activation and hair follicle regeneration. In cashmere goats, EGFR signaling may play a central role in regulating SHFSCs proliferation and migration, laying the foundation for cashmere fiber formation. EGFR is primarily expressed in the ORS and DP of SHFs, and immunohistochemical analysis showing ADAM17 localization in these regions further supports its role in epithelial-mesenchymal interactions, which are central to SHFs cycle regulation. Previous studies have shown that EGFR-mediated signaling not only promotes HFSC amplification but also maintains the dynamic balance of the hair follicle microenvironment by regulating inflammation and ECM remodeling (43, 44).

As an antagonist of the Wnt signaling pathway, SFRP1 was highly expressed in the ORS and HM during the anagen phase, indicating that it may finely regulate Wnt signaling to balance SHFSCs activation and differentiation. The Wnt signaling pathway plays a dual role in hair follicle regulation: it can activate β -catenin to promote stem cell proliferation or, under certain conditions, inhibit differentiation to maintain stem cell quiescence (45, 46). The high expression of SFRP1

may act as a “regulator,” partially suppressing Wnt signaling to prevent excessive stem cell differentiation, thereby maintaining the normal hair follicle growth cycle (47). This mechanism is particularly important in cashmere goats, as the periodic renewal of SHFs directly determines the growth quality and length of cashmere fibers. Interestingly, the role of SFRP1 in cashmere goat hair follicles differs from studies in mouse models, where SFRP1 was reported to inhibit Wnt-driven hair follicle activation (48). This discrepancy may reflect species-specific differences in hair follicle biology or the unique developmental requirements of SHFs in cashmere goats. Further research into the regulation of Wnt signaling by SFRP1 across species will help elucidate its universal and specific roles in hair follicle cycle regulation.

PPP1CA, a phosphatase involved in cell cycle regulation, was prominently expressed in the ORS, DP, and HM during the anagen phase. This expression pattern suggests a role in maintaining SHFSCs proliferation and differentiation, consistent with its role in promoting glioblastoma growth (49). PPP1CA facilitates the G1/S transition of the cell cycle by dephosphorylating retinoblastoma protein, thereby accelerating cell proliferation (50). Additionally, PPP1CA may regulate the activity of cell cycle-related proteins such as c-Myc (51, 52), supporting the rapid proliferation of HMCs to meet the demands of rapid SHFs growth during the anagen phase. Given the DP's role as the “signaling center” of hair follicle growth, the high expression of PPP1CA in the DPCs may promote SHFSCs activation and differentiation by regulating the secretion of signaling molecules.

In conclusion, ADAM17, SFRP1, and PPP1CA likely work in concert to regulate the telogen-to-anagen transition and promote cashmere fiber production. ADAM17 activates SHFSCs through the EGFR signaling pathway; SFRP1 balances Wnt signaling to regulate stem cell activation and promoting the differentiation of HMCs; and PPP1CA promotes SHF cell proliferation by modulating phosphorylation states. The precise spatial and temporal regulation of these proteins collectively drives the periodic renewal of cashmere goat SHFs.

Despite the significant findings of this study, several limitations remain. First, the small sample size may limit the generalizability of the results. Second, although the expression patterns of ADAM17, SFRP1, and PPP1CA were characterized, their functions were not validated through gene knockout or overexpression experiments. Future studies should focus on *in vivo* and *in vitro* functional analyses to elucidate the specific mechanisms of these molecules. Additionally, exploring the upstream regulators and downstream targets of these proteins will further enhance our understanding of their roles in hair follicle biology.

5 Conclusion

This study, through the integration of morphological analysis, proteomics, and experimental validation, uncovered dynamic changes during the telogen-to-anagen transition in cashmere goat SHFs. Furthermore, it established a connection between key DEPs—ADAM17, SFRP1, and PPP1CA—and the regulation of the SHF cycle. These findings not only deepen our understanding of the molecular mechanisms regulating the SHF cycle but also highlight potential molecular targets for enhancing cashmere fiber quality.

Data availability statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD060635.

Ethics statement

The animal study was approved by Institutional Animal Care and Use Ethics Committee of Inner Mongolia University. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

XH: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. GG: Conceptualization, Project administration, Supervision, Writing – review & editing. NS: Investigation, Methodology, Resources, Validation, Writing – review & editing. BD: Data curation, Formal analysis, Software, Visualization, Writing – review & editing. LR: Investigation, Validation, Writing – review & editing. HB: Formal analysis, Methodology, Software, Writing – review & editing. JiaL: Data curation, Investigation, Writing – review & editing. JieL: Validation, Visualization, Writing – review & editing. HZ: Data curation, Resources, Writing – review & editing. DL: Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China, grant no. U23A20226 and the Science and Technology Plan Project of

Inner Mongolia Autonomous Region of China, grant no. 2023KYPT0014.

Acknowledgments

We gratefully acknowledge the members of the laboratory for their suggestions and critical reading of the 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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.2025.1542682/full#supplementary-material>

References

- Dai B, Hao F, Xu T, Zhu B, Ren LQ, Han XY, et al. Thymosin B4 identified by transcriptomic analysis from Hf Anagen to Telogen promotes proliferation of Shf-Dpcs in albas cashmere goat. *Int J Mol Sci.* (2020) 21:268. doi: 10.3390/ijms21072268
- Hu XM, Li ZX, Zhang DY, Yang YC, Fu SA, Zhang ZQ, et al. A systematic summary of survival and death Signalling during the life of hair follicle stem cells. *Stem Cell Res Ther.* (2021) 12:453. doi: 10.1186/s13287-021-02527-y
- Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic Miniorgan. *Curr Biol.* (2009) 19:R132–42. doi: 10.1016/j.cub.2008.12.005
- Stenn KS, Paus R. Controls of hair follicle cycling. *Physiol Rev.* (2001) 81:449–94. doi: 10.1152/physrev.2001.81.1.449
- Lo Celso C, Prowse DM, Watt FM. Transient activation of Beta-catenin Signalling in adult mouse epidermis is sufficient to induce new hair follicles but continuous activation is required to maintain hair follicle Tumours. *Development.* (2004) 131:1787–99. doi: 10.1242/dev.01052
- Deschene ER, Myung P, Rompolas P, Zito G, Sun TY, Taketo MM, et al. B-catenin activation regulates tissue growth non-cell autonomously in the hair stem cell niche. *Science.* (2014) 343:1353–6. doi: 10.1126/science.1248373
- Aoi N, Inoue K, Chikanishi T, Fujiki R, Yamamoto H, Kato H, et al. 1 α ,25-Dihydroxyvitamin D3 modulates the hair-inductive capacity of dermal papilla cells: therapeutic potential for hair regeneration. *Stem Cells Transl Med.* (2012) 1:615–26. doi: 10.5966/sctm.2012-0032
- Hawkshaw NJ, Hardman JA, Alam M, Jimenez F, Paus R. Deciphering the molecular morphology of the human hair cycle: Wnt Signalling during the Telogen-to-anagen transformation. *Br J Dermatol.* (2020) 182:1184–93. doi: 10.1111/bjd.18356
- Vauclair S, Nicolas M, Barrandon Y, Radtke F. Notch1 is essential for postnatal hair follicle development and homeostasis. *Dev Biol.* (2005) 284:184–93. doi: 10.1016/j.ydbio.2005.05.018
- Estrach S, Ambler CA, Lo Celso C, Hozumi K, Watt FM. Jagged 1 is a Beta-catenin target gene required for ectopic hair follicle formation in adult epidermis. *Development.* (2006) 133:4427–38. doi: 10.1242/dev.02644
- Estrach S, Cordes R, Hozumi K, Gossler A, Watt FM. Role of the notch ligand Delta1 in embryonic and adult mouse epidermis. *J Invest Dermatol.* (2008) 128:825–32. doi: 10.1038/sj.jid.5701113
- Kang JI, Kim MK, Lee JH, Jeon YJ, Hwang EK, Koh YS, et al. Undariopsis Peterseniana promotes hair growth by the activation of Wnt/B-catenin and Erk pathways. *Mar Drugs.* (2017) 15, 130. doi: 10.3390/md15050130
- Kang JI, Yoon HS, Kim SM, Park JE, Hyun YJ, Ko A, et al. Mackerel-derived fermented fish oil promotes hair growth by Anagen-stimulating pathways. *Int J Mol Sci.* (2018) 19:770. doi: 10.3390/ijms19092770
- Tajan M, Paccoud R, Branka S, Edouard T, Yart A. The Rasopathy family: consequences of germline activation of the Ras/Mapk pathway. *Endocr Rev.* (2018) 39:676–700. doi: 10.1210/er.2017-00232

15. Choi S, Zhang B, Ma S, Gonzalez-Celeiro M, Stein D, Jin X, et al. Corticosterone inhibits Gas6 to govern hair follicle stem-cell quiescence. *Nature*. (2021) 592:428–32. doi: 10.1038/s41586-021-03417-2
16. Harel S, Higgins CA, Cerise JE, Dai Z, Chen JC, Clynes R, et al. Pharmacologic inhibition of Jak-Stat signaling promotes hair growth. *Sci Adv*. (2015) 1:e1500973. doi: 10.1126/sciadv.1500973
17. Wang ECE, Dai Z, Ferrante AW, Drake CG, Christiano AM. A subset of Trem2(+) dermal macrophages secretes Oncostatin M to maintain hair follicle stem cell quiescence and inhibit hair growth. *Cell Stem Cell*. (2019) 24:654–69.e6. doi: 10.1016/j.stem.2019.01.011
18. Elbediwy A, Vincent-Mistiaen ZI, Spencer-Dene B, Stone RK, Boeing S, Wculek SK, et al. Integrin Signalling regulates yap and Taz to control skin homeostasis. *Development*. (2016) 143:1674–87. doi: 10.1242/dev.133728
19. Chen Y, Fan Z, Wang X, Mo M, Zeng SB, Xu RH, et al. Pi3k/Akt signaling pathway is essential for De novo hair follicle regeneration. *Stem Cell Res Ther*. (2020) 11:144. doi: 10.1186/s13287-020-01650-6
20. Wang X, Chen H, Tian R, Zhang Y, Drutskaya MS, Wang C, et al. Macrophages induce Akt/B-catenin-dependent Lgr5(+) stem cell activation and hair follicle regeneration through Tnf. *Nat Commun*. (2017) 8:14091. doi: 10.1038/ncomms14091
21. Geyfman M, Plikus MV, Treffeisen E, Andersen B, Paus R. Resting no more: re-defining Telogen, the maintenance stage of the hair growth cycle. *Biol Rev Camb Philos Soc*. (2015) 90:1179–96. doi: 10.1111/brv.12151
22. Millar SE. Molecular mechanisms regulating hair follicle development. *J Invest Dermatol*. (2002) 118:216–25. doi: 10.1046/j.0022-202x.2001.01670.x
23. Niu Y, Wang Y, Chen H, Liu X, Liu J. Overview of the circadian clock in the hair follicle cycle. *Biomol Ther*. (2023) 13:1068. doi: 10.3390/biom13071068
24. Wang E, Harel S, Christiano AM. Jak-Stat signaling jump starts the hair cycle. *J Invest Dermatol*. (2016) 136:2131–2. doi: 10.1016/j.jid.2016.08.029
25. Aggarwal S, Talukdar NC, Yadav AK. Advances in higher order multiplexing techniques in proteomics. *J Proteome Res*. (2019) 18:2360–9. doi: 10.1021/acs.jproteome.9b00228
26. Ji S, Zhu Z, Sun X, Fu X. Functional hair follicle regeneration: an updated review. *Signal Transduct Target Ther*. (2021) 6:66. doi: 10.1038/s41392-020-00441-y
27. Kim H, Jang Y, Kim EH, Jang H, Cho H, Han G, et al. Potential of colostrum-derived exosomes for promoting hair regeneration through the transition from Telogen to Anagen phase. *Front Cell Dev Biol*. (2022) 10:815205. doi: 10.3389/fcell.2022.815205
28. Li G, Tang X, Zhang S, Jin M, Wang M, Deng Z, et al. Sirt7 activates quiescent hair follicle stem cells to ensure hair growth in mice. *EMBO J*. (2020) 39:e104365. doi: 10.15252/embj.2019104365
29. Panteleyev AA. Functional anatomy of the hair follicle: the secondary hair germ. *Exp Dermatol*. (2018) 27:701–20. doi: 10.1111/exd.13666
30. Sennett R, Rendl M. Mesenchymal-epithelial interactions during hair follicle morphogenesis and cycling. *Semin Cell Dev Biol*. (2012) 23:917–27. doi: 10.1016/j.semcdb.2012.08.011
31. Zhou Y, Seo J, Tu S, Nanmo A, Kageyama T, Fukuda J. Exosomes for hair growth and regeneration. *J Biosci Bioeng*. (2024) 137:1–8. doi: 10.1016/j.jbiosc.2023.11.001
32. Li KN, Tumber T. Hair follicle stem cells as a skin-organizing signaling center during adult homeostasis. *EMBO J*. (2021) 40:e107135. doi: 10.15252/embj.2020107135
33. Zhang B, Chen T. Local and systemic mechanisms that control the hair follicle stem cell niche. *Nat Rev Mol Cell Biol*. (2024) 25:87–100. doi: 10.1038/s41580-023-00662-3
34. Peschon JJ, Slack JL, Reddy P, Stocking KL, Sunnarborg SW, Lee DC, et al. An essential role for Ectodomain shedding in mammalian development. *Science*. (1998) 282:1281–4. doi: 10.1126/science.282.5392.1281
35. Freed DM, Bessman NJ, Kiyatkin A, Salazar-Cavazos E, Byrne PO, Moore JO, et al. Egfr ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell*. (2017) 171:683–95.e18. doi: 10.1016/j.cell.2017.09.017
36. Watanabe T, Shintani A, Nakata M, Shing Y, Folkman J, Igarashi K, et al. Recombinant human Betacellulin. Molecular structure, biological activities, and receptor interaction. *J Biol Chem*. (1994) 269:9966–73. doi: 10.1016/s0021-9258(17)36977-6
37. Kim J, Jahng WJ, di D, Lee JS, Jhaveri R, Rubin MA, et al. The phosphoinositide kinase Pikfyve mediates epidermal growth factor receptor trafficking to the nucleus. *Cancer Res*. (2007) 67:9229–37. doi: 10.1158/0008-5472.Can-07-1333
38. Zhang H, Nan W, Wang S, Zhang T, Si H, Yang F, et al. Epidermal growth factor promotes proliferation and migration of follicular outer root sheath cells via Wnt/B-catenin signaling. *Cell Physiol Biochem*. (2016) 39:360–70. doi: 10.1159/000445630
39. Gentile P, Garcovich S. Advances in regenerative stem cell therapy in androgenic alopecia and hair loss: Wnt pathway, growth-factor, and mesenchymal stem cell signaling impact analysis on cell growth and hair follicle development. *Cells*. (2019) 8:466. doi: 10.3390/cells8050466
40. Kumtorrut C, Yamauchi T, Koike S, Aiba S, Yamasaki K. Androgens modulate keratinocyte differentiation indirectly through enhancing growth factor production from dermal fibroblasts. *J Dermatol Sci*. (2019) 93:150–8. doi: 10.1016/j.jdermsci.2019.01.007
41. Sugawara K, Schneider MR, Dahlhoff M, Kloepper JE, Paus R. Cutaneous consequences of inhibiting Egf receptor signaling in vivo: Normal hair follicle development, but retarded hair cycle induction and inhibition of adipocyte growth in Egfr(Wa5) mice. *J Dermatol Sci*. (2010) 57:155–61. doi: 10.1016/j.jdermsci.2009.12.004
42. Choi N, Kim WS, Oh SH, Sung JH. Epiregulin promotes hair growth via Egfr-mediated epidermal and Erbb4-mediated dermal stimulation. *Cell Prolif*. (2020) 53:e12881. doi: 10.1111/cpr.12881
43. Saad MI, Jenkins BJ. The protease Adam17 at the crossroads of disease: revisiting its significance in inflammation, Cancer, and beyond. *FEBS J*. (2024) 291:10–24. doi: 10.1111/febs.16923
44. Scheller J, Chalaris A, Garbers C, Rose-John S. Adam17: a molecular switch to control inflammation and tissue regeneration. *Trends Immunol*. (2011) 32:380–7. doi: 10.1016/j.it.2011.05.005
45. Wang X, Liu Y, He J, Wang J, Chen X, Yang R. Regulation of signaling pathways in hair follicle stem cells. *Burns Trauma*. (2022) 10:10. doi: 10.1093/burnst/tkac022
46. Lien WH, Polak L, Lin M, Lay K, Zheng D, Fuchs E. In vivo transcriptional governance of hair follicle stem cells by canonical Wnt regulators. *Nat Cell Biol*. (2014) 16:179–90. doi: 10.1038/ncb2903
47. Hawkshaw NJ, Hardman JA, Haslam IS, Shahmalak A, Gilhar A, Lim X, et al. Identifying novel strategies for treating human hair loss disorders: cyclosporine a suppresses the Wnt inhibitor, Sfrp1, in the dermal papilla of human scalp hair follicles. *PLoS Biol*. (2018) 16:e2003705. doi: 10.1371/journal.pbio.2003705
48. Sunkara RR, Mehta D, Sarate RM, Waghmare SK. Bmp-Akt-Gsk3 β signaling restores hair follicle stem cells decrease associated with loss of Sfrp1. *Stem Cells*. (2022) 40:802–17. doi: 10.1093/stmcls/sxac041
49. Ladha J, Donakonda S, Agrawal S, Thota B, Srividya MR, Sridevi S, et al. Glioblastoma-specific protein interaction network identifies Pp1a and Csk21 as connecting molecules between cell cycle-associated genes. *Cancer Res*. (2010) 70:6437–47. doi: 10.1158/0008-5472.Can-10-0819
50. Barnoud T, Donniger H, Clark GJ. Ras Regulates Rb Via Nore1a. *J Biol Chem*. (2016) 291:3114–23. doi: 10.1074/jbc.M115.697557
51. Dingar D, Tu WB, Resettec D, Lourenco C, Tamachi A, de J, et al. Myc Dephosphorylation by the Pp1/Pnnts phosphatase complex regulates chromatin binding and protein stability. *Nat Commun*. (2018) 9:3502. doi: 10.1038/s41467-018-05660-0
52. Xing J, Zhang Z, Mao H, Schnellmann RG, Zhuang S. Src regulates cell cycle protein expression and renal epithelial cell proliferation via Pi3k/Akt signaling-dependent and -independent mechanisms. *Am J Physiol Renal Physiol*. (2008) 295:F145–52. doi: 10.1152/ajprenal.00092.2008



OPEN ACCESS

EDITED BY

Zhihong Liu,
Inner Mongolia Agricultural University, China

REVIEWED BY

Ningbo Chen,
Northwest A&F University, China
Zeyang Wang,
Shenyang Agricultural University, China

*CORRESPONDENCE

Haoyuan Han,
✉ hanhaoyuan@126.com

RECEIVED 20 December 2024

ACCEPTED 26 February 2025

PUBLISHED 24 March 2025

CITATION

Quan K, Shi H, Wei C, Li J, Liu K, Wang H, Sun W
and Han H (2025) Genetic diversity,
reproductive performance, and genetic
enhancement strategies in Huang-Huai goats.
Front. Genet. 16:1549051.
doi: 10.3389/fgene.2025.1549051

COPYRIGHT

© 2025 Quan, Shi, Wei, Li, Liu, Wang, 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 diversity, reproductive performance, and genetic enhancement strategies in Huang-Huai goats

Kai Quan¹, Huibin Shi¹, Caihong Wei², Jun Li¹, Kun Liu¹,
Huihua Wang², Wei Sun³ and Haoyuan Han^{1*}

¹College of Animal Science and Technology, Henan University of Animal Husbandry and Economy, Zhengzhou, Henan, China, ²Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, ³College of Animal Science and Technology, Yangzhou University, Yangzhou, China

The Huang-huai goat, indigenous to China's Huang-Huai Plain, is celebrated for its exceptional reproductive capacity, succulent meat, and superior leather qualities. The Huang-huai goat's reproductive characteristics, genetic diversity, and the genetic underpinnings of intersexuality, aiming to inform conservation efforts and genetic resource management. Our study at the *Huang-huai Goat Science and Technology R&D Center* monitored 600–800 female goats and 16–24 male goats from June 2020 to May 2022, adhering to NIH guidelines and with ethical approval from *Henan University of Animal Husbandry and Economics*. Our findings indicate that these goats exhibit a year-round estrus cycle averaging 19–23 days, a gestation period of 146–150 days, and an average litter size of 2.74, with an annual reproduction rate of 418.96% and a weaning survival rate of 94.75%. Transcriptome sequencing identified eleven candidate genes associated with multiple offspring, including *PTX3*, *MMP13*, and *NR4A1*, which play roles in organ development and hormonal regulation. SNP analysis revealed specific genotypes in *GJB6* and *PRKAA1* linked to higher lambing numbers, offering molecular markers for selective breeding. The study also highlighted the role of the Polled Intersex Syndrome (PIS) locus in causing both hornless and intersexual traits, emphasizing the importance of genetic screening for maintaining breed health and productivity. The genetic resources of the Huang-huai goat, recognized as a national geographical indication product, are vital for the livestock industry and require strategic conservation for sustainable development. This review highlights the importance of preserving and utilizing the genetic resources of the Huang-huai goat to enhance its contribution to agriculture.

KEYWORDS

Huang-huai goat, hereditary, reproduction, intersexuality, PIS

1 Introduction

Nestled between 33° and 35.5°N latitude and 113° and 119°E longitude, the Huang-Huai Plain boasts a climate of humidity and ample sunlight, alongside an abundance of natural resources that have historically established it as the cradle of ancient Chinese civilization (Zhou et al., 2014; Hu et al., 2021). The region's advantageous irrigation systems and substantial agricultural byproducts, particularly crop straw, make it an ideal hub for

industrial-scale agriculture and animal husbandry, and it has been duly recognized as a pivotal grain-producing area in China (Xiao et al., 2022). This fertile landscape is especially hospitable to goat farming.

The Huang-Huai goat, also known as the Huai goat, derives its name from the Huang-Huai Plain where it has its origin (Zhou et al., 2020). There are over 10 million of these goats dispersed across the eastern part of Henan, the northern regions of Anhui, and the northern areas of Jiangsu (Li et al., 2022). These medium-sized goats possess short white hair and pink skin. They can be classified into horned (accounting for 43.03%) and hornless (accounting for 56.97%) varieties (Hua et al., 2008). On average, adult male goats weigh 49.08 kg, while adult female goats weigh 37.75 kg (Pan et al., 2011). Since 2000, excessive crossbreeding of Huang-huai goats with Boer goats has led to the degradation of their germplasm advantages—including high fertility, superior meat quality, and premium leather traits—resulting in diminished reproductive efficiency and compromised market reputation (Han et al., 2022). Unrestricted hybridization has resulted in a deterioration in reproductive performance, as well as in the quality of meat and hides (Yang et al., 2021). In particular, the high occurrence rate of intersexuality has hindered the selection and propagation of the hornless trait, posing a threat to the future of this breed (Han et al., 2022; Yang et al., 2021).

To enhance the conservation and efficient use of Huang-Huai goat germplasm resources, the expert team of the *Henan agricultural research system of sheep and Goat Industry Technology (HARS-22–15)*, relying on the *Shenqiu Huai Goat Technology R&D Center*, has conducted a comprehensive and systematic assessment of its production performance over a period of 5 years. This review summarizes the reproductive characteristics and progress in genetic improvement of the Huang-Huai goat.

2 Genetic diversity and maternal origin of Huang-huai goats

2.1 Genetic diversity and its significance

Genetic diversity is a crucial indicator for the conservation of germplasm resources. Studies have revealed that the Huang-Huai goat, along with several other local breeds such as the Funiu white goat and Yaoshan white goat, exhibit relatively high genetic polymorphism. This indicates that these breeds possess abundant genetic resources and strong adaptability to the environment.

2.2 Maternal origin analysis

Maternal origin analysis has demonstrated that these breeds mainly originated from the Markhor. There are two maternal lineages, namely lineage A and lineage B, which reveals the complex evolutionary history and genetic background of the Huang-Huai goat.

2.3 Mitochondrial DNA research

Research on mitochondrial DNA (mtDNA) has provided direct evidence for the maternal origin of the Huang-Huai goat. Studies on

the mitochondrial D-loop region of domestic goats in the central region have shown that the maternal origin of these breeds is the Markhor. Moreover, the majority of individuals are from lineage A, while a small number are from lineage B, further confirming the hypothesis of multiple maternal origins of the Huang-Huai goat.

3 Reproductive performance of Huang-huai goats

The research on the reproductive performance of Huang-Huai goats was carried out at the *Huang-Huai Goat Science and Technology R&D Center*, which is dedicated to the study, demonstration, and promotion of this particular breed. From June 2020 to May 2022, the center accommodated 600–800 female goats and 16–24 male goats of the Huang-Huai breed. All the animals were taken care of in strict accordance with the guidelines stipulated by the National Institutes of Health regarding the care and use of experimental animals. Moreover, all the procedures and experiments were duly approved by the *Animal Care and Use Committee of Henan University of Animal Husbandry and Economics*.

3.1 Study design and data collection

A reproduction test was carried out on Huang-Huai goats. Daily observations and data collection were made concerning estrus performance, breeding status, heat duration, and male goat semen quality. Data filtering was implemented to exclude any values that deviated beyond the mean $\pm$ standard deviation range for each trait. Statistical analysis was then performed on various parameters, including the first estrus period, estrus duration, breeding cycle, average litter size, lamb weaning survival rate, average number of weaned lambs per female per year, and male goat semen quality.

3.2 Puberty and reproductive characteristics

Male lambs display signs of puberty through behaviors such as sniffing the vulva of female goats, chasing them upon detecting estrus (sexual olfactory reflex), barking after olfactory perception, penile erection, mounting, and ejaculation (Quan et al., 2020). The average age of puberty for male lambs ranges from 75 to 83 days, with an average weight of 15–23 kg. Female lambs exhibit symptoms of puberty like mental excitement, loss of appetite, tail erection, and vulvar swelling accompanied by secretions, although they do not permit male goats to mount them. The average age and weight at puberty for female goats are 63–84 days and 14–22 kg respectively. They reach sexual maturity at 4–5 months of age, with a weight of 18–25 kg. Huang-Huai goats show a relatively higher proportion of estrus in autumn and winter, yet they can experience estrus throughout the year, with an estrus cycle spanning from 18.3 to 22.8 days. The gestation period is 146–150 days, and the average lambing interval is 238.34 days. The average litter size is 2.74, and the annual reproduction rate amounts to 418.96% (Table 1). The weaning survival rate is 94.75%, with an average of 3.97 weaned lambs per female per year (Table 2).

TABLE 1 The estrus and lambing of female goats in the month from June 2020 to May 2022.

Months	Female goats (n)	Number of estrus goats (n)	Number of delivery female goats (n)	Number of lambs born (n)	Number of weaned lambs (n)
January	1,242	165	152	433	421
February	1,233	144	187	537	442
March	1,238	130	205	565	426
April	1,267	143	196	462	468
May	1,241	167	174	488	457
June	1,302	168	151	435	454
July	1,271	191	133	361	421
August	1,236	166	121	302	386
September	1,265	205	136	385	320
October	1,283	224	154	443	371
November	1,303	219	149	402	413
December	1,242	191	172	467	424
Total	-	2,113	1930	5,280	5,003

The data are the sum of 2 years, the number of female goats was the statistics at the end of last month, and the number of total female goats was average of 2 years.

TABLE 2 Statistics of different litter size.

Item	Number of lambing female goats (n)	Number of lambs (n)	Weaning lamb number (n)	Lamb weaning survival rate (%)
Single	43	43	43	100.00
Double	476	952	950	99.79
Triple	1,364	4,092	3,871	94.60
Quadruple	42	168	124	73.81
Quintuple	5	25	15	60.00
Total	1930	5,280	5,003	94.75

3.3 Semen quality and reproduction seasonality

Huang-Huai male goats are capable of reproducing throughout the year. Their semen quality reaches its peak in autumn and winter and deteriorates in summer (Table 3). The average ejaculation volume is 1.01 ± 0.11 mL, with an average sperm motility of $84.49\% \pm 2.26\%$ and a sperm concentration of $(21.52 \pm 6.86) \times 10^8/\text{mL}$. The defective sperm rate is $12.16\% \pm 2.31\%$, and the semen is usually milky white.

3.4 Reproductive performance and breeding efficiency

Reproductive performance is a crucial index for evaluating goat production levels, especially in large-scale farming systems. In comparison with other Chinese local breeds and introduced breeds like Boer goats (Wang et al., 2020), Huang-Huai goats reach puberty and sexual maturity at an earlier stage. They can

reproduce throughout the year with minimal seasonal impact, attaining an average of 1.57 pregnancies per year, and each female can wean approximately 3.97 lambs annually.

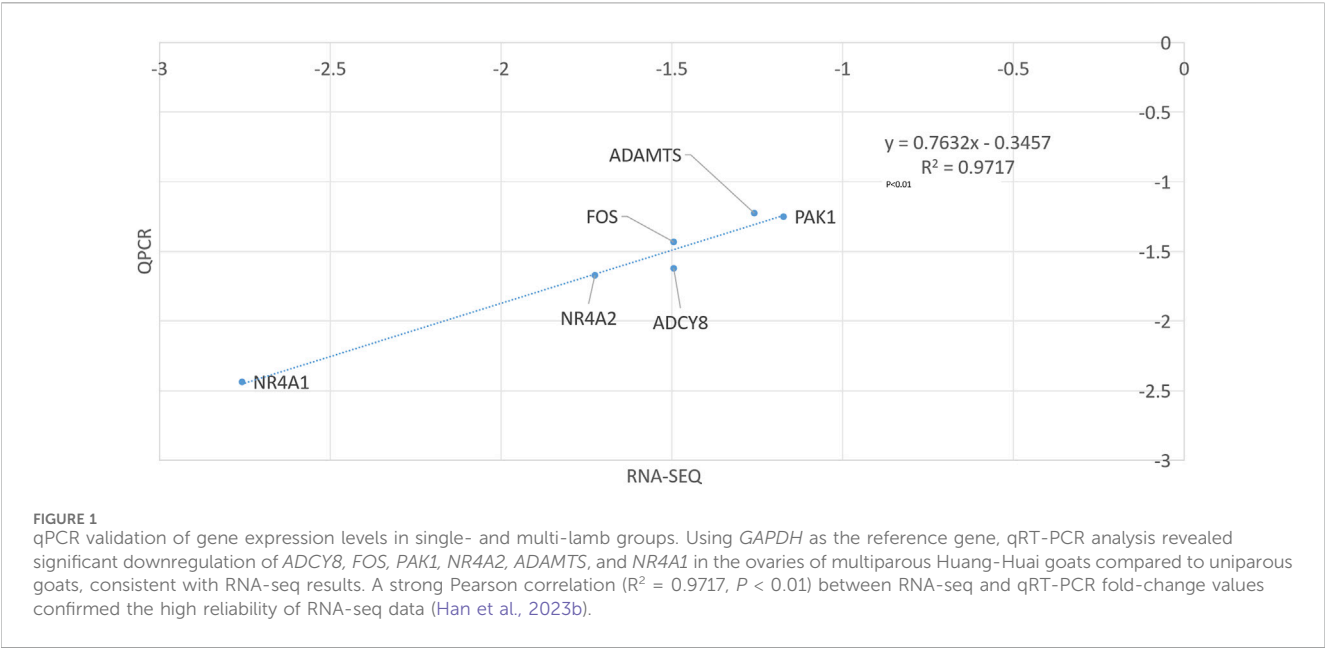
The early sexual maturity and high reproductive performance of Huang-Huai goats confer significant benefits upon breeders. It enables early breeding and shorter generation intervals, which can boost productivity and potentially augment profits (Chuet al., 2011). Although the breed's inherent characteristics contribute to its early reproductive development, optimal management and a balanced diet are indispensable for maintaining the peak reproductive performance in goat farming.

4 Candidate genes for multiple lambs in Huang-huai goats

To understand the genetic basis of multiple lambs in Huang-Huai goats, 3-year-old healthy female goats with similar body conditions were selected and grouped based on their reproductive records: a single-lamb group (24 females, averaging one lamb per

TABLE 3 Semen quality of male goats in different seasons.

Season	Number of male goats(n)	Ejaculate volumes (mL)	Sperm motility rate (%)	Sperm concentration (10 ⁹ /mL)	Defective sperm rate (%)
Spring	26	1.01 ± 0.11	81.12 ± 3.22	21.53 ± 7.29	13.42 ± 2.92
Summer	20	0.88 ± 0.17	80.73 ± 3.59	16.59 ± 8.18	14.34 ± 2.67
Autumn	36	1.23 ± 0.08	89.14 ± 1.21	25.14 ± 6.34	10.12 ± 1.31
Winter	31	1.19 ± 0.12	84.33 ± 1.83	22.83 ± 5.61	11.42 ± 2.71
Average	-	1.01 ± 0.11	84.49 ± 2.26	21.52 ± 6.86	12.16 ± 2.31



birth) and a multi-lamb group (24 females, averaging over two lambs per birth). Transcriptome sequencing was used to build mRNA libraries for both groups to identify differentially expressed genes (DEGs), followed by functional enrichment analysis and verification with real-time fluorescence quantitative PCR. Blood genomic DNA from goats with different reproductive performances was extracted. Based on preliminary transcriptome analysis and literature review, *GJB6* (gap junction protein beta 6) and *PRKAA1* (AMP-activated protein kinase catalytic subunit alpha-1) were prioritized for SNP analysis due to their established roles in follicular development and metabolic regulation of reproduction in other ruminants (Domínguez-Ruiz et al., 2024; Wang et al., 2020; Xiao et al., 2020). Candidate genes related to multiple lambs were SNP scanned and genotyped, and the correlation with high lambing frequencies was analyzed. Primers for *GJB6* (forward: 5'-CAG GTGCTGGACTTCATCCT-3'; reverse: 5'-TGGCAATGTCAC AGAGGACA-3') and *PRKAA1* (forward: 5'-GCTGGACCTCAA CCTGATGA-3'; reverse: 5'-AGCCACAGGGTCTTCATGGT-3') were designed using Primer-BLAST (NCBI).

Eleven candidate genes related to lambing number in Huang-Huai goats were found by comparing the transcriptomes of the two groups (Gawat et al., 2023): *PTX3*, *MMP13*, *PAK1*, *ADAMTS1*, *COL1A2*, *CCN1*, *SLC4A10*, *FOS*, *NR4A1*, *NR4A2*, and *ADCY8*. These genes are involved in animal organ development and

hormone secretion in the endocrine system (Yao et al., 2023), and qPCR validation confirmed differential expression patterns ($P < 0.05$) between single- and multi-lamb groups (Figure 1).

Specific primers were designed for the *GJB6* and *PRKAA1* genes in Huang-Huai goats. After sequencing and genotyping, two synonymous mutations in *GJB6* and three SNP sites in the non-coding sequence of *PRKAA1* were detected. The TC genotype of g.33691489 T>C and the AT genotype of g.33693395 A>T in *PRKAA1* were associated with a higher number of lambs (Table 4). Haplotype analysis further revealed that the TCA haplotype (comprising SNPs g.33691489 T>C, g.33693100 T>C, and g.33693395 A>T) was significantly associated with increased prolificacy ($\chi^2 = 8.24$, $P = 0.004$), aligning with linkage disequilibrium patterns in the *PRKAA1* locus.

Transcriptome sequencing helps find DEGs among different phenotypes. Some genes related to follicular development, ovulation, and steroid hormone synthesis are expressed in ovarian tissues of Huang-Huai goats (Ceccobelli et al., 2023). For example, *MMP13* is important for follicle formation and ovulation, and *ADAMTS1* promotes follicular development and maturation. *NR4A1* and *NR4A2* are involved in the ARH response related to reproductive function and may regulate lambing (McNulty et al., 2012).

SNPs are related to gene expression and individual trait differences. In Huang-Huai goats, the TC genotype at

TABLE 4 The association analysis between SNPs of *GJB6/PRKAA1* gene and lambing numbers of Huai goats (average $\pm$ standard deviation).

Gene	SNP	Genotype	No.	Lambing number
<i>GJB6</i>	g.50694819 C>T	CC	103	1.670 $\pm$ 0.103
		CT	16	1.711 $\pm$ 0.169
		TT	3	1.654 $\pm$ 0.371
	g.50694816 T>C	CC	8	1.747 $\pm$ 0.223
		CT	42	1.710 $\pm$ 0.165
		TT	72	1.578 $\pm$ 0.168
<i>PRKAA1</i>	g.33691489 T>C	TT	25	1.600 $\pm$ 0.139 ^{AB}
		TC	58	1.953 $\pm$ 0.125 ^A
		CC	39	1.350 $\pm$ 0.169 ^B
	g.33693100 T>C	TT	103	1.557 $\pm$ 0.134
		TC	16	1.738 $\pm$ 0.205
		CC	3	1.000 $\pm$ 0.400
	g.33693395 A>T	AA	34	1.492 $\pm$ 0.250 ^B
		AT	63	1.780 $\pm$ 0.151 ^A
		TT	25	1.321 $\pm$ 0.170 ^B

Haplotype frequencies were analyzed using PHASE, v2.1, and association tests were performed with PLINK v1.9. Different uppercase letters indicate extremely significant difference ($P < 0.01$), same letters or not marked indicate no significant difference ($P > 0.05$) (Han et al., 2023a).

g.33691489 T>C and the AT genotype at g.33693359 A>T in *PRKAA1* are significantly associated with a higher lambing number (Wang et al., 2019). Haplotype analysis shows that the TCA haplotype is also related to increased lambing. These loci can be potential molecular markers for selecting multiple lambing traits.

5 Intersex syndrome in Huang-huai goats

Six intersex Huang-Huai goats were slaughtered to study their reproductive organ anatomy (Han et al., 2022; Yang et al., 2021). HE staining was performed on ovarian samples to histologically evaluate tissue structure, identify pathological changes (e.g., testicular-like lesions), and confirm the presence of intersex-specific anomalies (e.g., mixed ovarian/testicular tissues), which are critical for diagnosing PIS. Blood samples from 18 intersex and 45 normal goats (12 females and 23 males) were collected. PCR was used to amplify the SRY gene and detect the PIS deletion.

Externally, intersex goats had a smaller vulva. Some had testes but no vulva and had underdeveloped genitals causing dysuria (Figure 2). Internally, sample A had ovarian-like gonads larger than normal ovaries with follicular structures and swollen female reproductive ducts. Sample B had one large and one small gonad with different connections to genital tracts and a nearby penis-like structure. Samples C, D, E, and F had uneven-surfaced gonads, with the left gonad having a spermatic cord and a degenerated uterus (Figure 3).

SRY amplification showed 18 intersex goats had XX sex chromosomes. Using Monteagudo’s and Simon’s methods (Monteagudo et al., 2008; Simon et al., 2020), it was found that

the 18 intersex goats had homozygous PIS deletions. Among 45 normal goats, 17 were PIS+/+ homozygous wild-type and 28 were PIS+/- heterozygous. Among 23 male goats, 8 were PIS+/+ and 15 (65.21%) were PIS+/- heterozygous. Among 22 female goats, 9 were PIS+/+ and 13 (59.09%) were PIS+/- heterozygous (Table 5).

In normal Huang-Huai goats, 62% carry PIS deletion, with 65.21% of male goats and 59.09% of female goats. Homozygous PIS deletion in offspring causes intersex traits. If carriers aren’t excluded, the intersex gene frequency will rise, reducing population reproduction and economic benefits. So, intersex detection in Huang-Huai goats is urgent. The study found that the PIS locus causes the hornless trait (autosomal dominant) and the intersexual trait (autosomal recessive) in Huang-Huai goats. The two traits are linked, with intersexuality usually in hornless goats and regulated by the PIS region (100 kb on chromosome 1q43, between *ATP1B* and *COP* genes, with a 11.7 kb fragment deletion causing the syndrome). The PIS region transcribes genes like *FOXL2*, *PISRT1*, and *PFOXic*.

6 Application of genetic improvement for Huang-huai goat

6.1 Enhancing reproductive performance through genetic selection

Our team has made significant strides in the genetic enhancement of the Huang-huai goat by optimizing the breeding core group for multiparous traits. Utilizing genomic tools, we have developed breeding chips that facilitate the identification of multiparous female goats. This technology has been applied



FIGURE 2
Observation of the external genitalia of intersexual Huang-huai goats. A-F represent external genitalia of intersexual Huang-huai goats. Intersexual Huang-huai goats have female reproductive organs, but with an enlarged, prominent clitoris. The vulva is smaller, and the clitoris resembles a penis. Some individuals may have seemingly male reproductive organs, but they are incompletely developed. Overall, their external genital morphology is intermediate between normal male and female goats (Yang et al., 2021).

across three conservation farms, leading to the establishment of a core breeding group comprising 1,200 goats. This targeted approach has accelerated the enrichment of multiparous genes, concentrating superior genetic traits within the core group and thereby enhancing the overall reproductive performance of the population.

6.2 Incorporating PIS testing to preserve genetic integrity

We have implemented PIS testing for male goats, a crucial step in maintaining the genetic health of the breed. This initiative involved screening 210 local breeding males, resulting in the exclusion of 134 individuals carrying the intersex gene. This

proactive measure has substantially reduced the prevalence of intersex traits among Huang-huai goats, safeguarding the genetic integrity of the breeding population and preventing the spread of deleterious genes, which is vital for the sustainable development and reproductive vigor of the breed.

7 Discussion

This review comprehensively elucidates the genetic diversity, reproductive traits, and genetic enhancement strategies of Huang-Huai goats, providing critical insights for breed conservation and sustainable utilization. The high genetic polymorphism and dual maternal lineages (A and B) underscore the breed's evolutionary

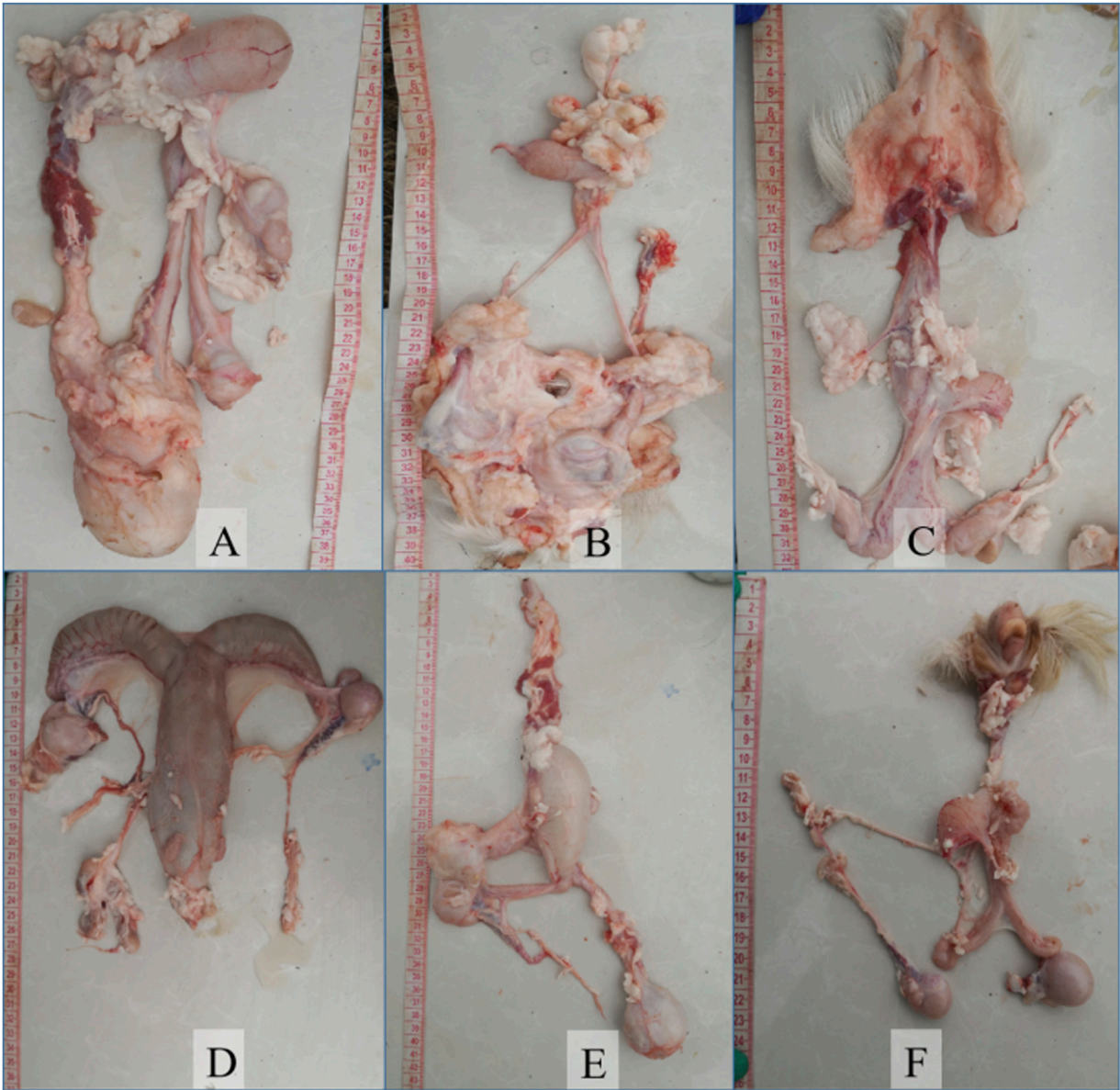


FIGURE 3
Observation of the internal genitalia of intersexual Huang-huai goats. A-F represent the gonads of intersexual Huang-huai goats. It showed significant morphological differences from typical female ovaries and are larger in size. Histologically, they show a mixed pattern, similar to ovarian tissue with testicular-like lesions or partially differentiated testicular structures. The reproductive system has well-defined female organs (uterus and vagina) but also underdeveloped male components (rudimentary ductal systems or incomplete spermatic cord formation), representing a transitional state between male and female reproductive organ development (Yang et al., 2021).

TABLE 5 Distinguishing types of PIS between normal and intersex Huang-huai goats.

Genotype individual	Number	Gender	Individual number
PIS ^{+/+}	17	♀	9
		♂	8
PIS ^{+/-}	28	♀	13
		♂	15

resilience, aligning with findings in other indigenous Chinese breeds (Peng et al., 2024). Notably, Huang-Huai goats exhibit exceptional reproductive performance, including year-round estrus, a high annual reproduction rate (418.96%), and robust weaning survival (94.75%), surpassing many local and introduced breeds (Han et al., 2022). The identification of candidate genes (*PTX3*, *MMP13*, *PRKAA1*) and SNPs associated with prolificacy offers novel molecular markers for selective breeding, complementing prior studies on fecundity-related genes in goats (Han et al., 2022). However, functional validation of these genes and their regulatory mechanisms in ovulation and hormone synthesis warrants further investigation.

The linkage between the PIS locus and intersexuality highlights a critical challenge in breeding hornless variants. The high carrier rate (62%) of PIS deletions emphasizes the urgency of genetic screening to mitigate intersex traits, as proposed in Polled goat management (Yang et al., 2021). While this study establishes a framework for PIS detection, broader validation across diverse populations is essential to ensure applicability. Additionally, environmental and epigenetic factors influencing reproductive performance remain unexplored, representing a limitation of the current dataset (Guang-Xin et al., 2018).

Future efforts should focus on expanding genomic databases, integrating multi-omics approaches to unravel gene-environment interactions, and refining breeding chips for precision selection (Liu et al., 2024). Collaborative initiatives to disseminate genetic screening protocols and core breeding strategies will enhance breed purity and productivity. This work lays the foundation for leveraging Huang-Huai goats' genetic potential, balancing conservation with agricultural innovation in China's livestock sector.

8 Conclusion

The exploration, conservation, and strategic utilization of the Huang-Huai goat germplasm resources are essential for the protection of biodiversity and the advancement of new, distinctive breeds (Zhang et al., 2024; Jin et al., 2020). This study has conducted a thorough examination of the genetic diversity, maternal lineage, and reproductive performance of the Huang-Huai goats. By concentrating on candidate genes associated with multiple births, we have developed a specialized breeding chip for the Huang-Huai goats and have selected a core breeding population. Additionally, to address the intersex syndrome observed in the Huang-Huai goats, a targeted screening protocol has been implemented to identify and exclude male goats carrying intersex genetic markers. These efforts have substantially enhanced the technical infrastructure for the conservation and efficient utilization of the Huang-Huai goat germplasm resources.

Author contributions

KQ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

HS: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft. CW: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft. JL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–review and editing. KL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft. HW: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft. WS: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–review and editing. HH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by Henan agricultural research system (HARS-22-15-S), and National natural science foundation of China (youth fund) (32402758), Henan province key research and promotion project (252102110068), and Program for Innovative Research Team (in Science and Technology) in University of Henan Province (25IRTSTHN027).

Acknowledgments

This study was supported by Henan Institute of Animal Husbandry and Economics of China, Shenqiu County Agricultural and Animal Husbandry Science and Technology Research and Development Center and team members.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

- Ceccobelli, S., Landi, V., Senczuk, G., Mastrangelo, S., Sardina, M. T., Ben-Jemaa, S., et al. (2023). A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds. *Genet. Sel. Evol. GSE* 55 (1), 24. doi:10.1186/s12711-023-00797-z
- Chu, M. X., Lu, L., Feng, T., Di, R., Cao, G. L., Wang, P. Q., et al. (2011). Polymorphism of bone morphogenetic protein 4 gene and its relationship with litter size of Jining Grey goats. *Mol. Biol. Rep.* 38 (7), 4315–4320. doi:10.1007/s11033-010-0556-6
- Domínguez-Ruiz, M., Murillo-Cuesta, S., Contreras, J., Cantero, M., Garrido, G., Martín-Bernardo, B., et al. (2024). A murine model for the del(GJB6-D13S1830) deletion recapitulating the phenotype of human DFNB1 hearing impairment: generation and functional and histopathological study. *BMC genomics* 25 (1), 359. doi:10.1186/s12864-024-10289-z
- Gawat, M., Boland, M., Singh, J., and Kaur, L. (2023). Goat meat: production and quality attributes. *Foods* 12 (16), 3130. doi:10.3390/foods12163130
- Guang-Xin, E., Zhao, Y. J., Chen, L. P., Ma, Y. H., Chu, M. X., Li, X. L., et al. (2018). Genetic diversity of the Chinese goat in the littoral zone of the Yangtze River as assessed by microsatellite and mtDNA. *Ecol. Evol.* 8 (10), 5111–5123. doi:10.1002/ece3.4100
- Han, H., Yang, S., Li, J., Zhao, J., Wei, H., Ha, S., et al. (2022). Intersex goats show different gene expression levels in the hypothalamus and pituitary compared with non-intersex goats based on RNA-Seq. *Veterinary Med. Sci.* 8 (1), 367–376. doi:10.1002/vms3.672
- Han, H., Li, T., Li, S., Song, X., Li, J., Ha, S., et al. (2023a). Polymorphism of *GJB6* and *PRKAA1* genes and their association with litter size in Huishan goats. *Acta Agriculturae Borealis - Sinica*. 38 (04): 225–232. (in Chinese).
- Han, H., Li, S., Yang, R., Li, M., Li, J., Ha, S., et al. (2023b). Identification of key candidate genes for high fertility in Huishan goats based on transcriptome sequencing. *Acta Veterinaria et Zootechnica Sinica*. 54 (12): 5077–5090. (in Chinese).
- Hu, Z., Wu, Z., Zhang, Y., Li, Q., Islam, A. R. M. T., and Pan, C. (2021). Risk assessment of drought disaster in summer maize cultivated areas of the Huang-Huai-Hai plain, eastern China. *Environ. Monit. Assess.* 193 (7), 441. doi:10.1007/s10661-021-09224-6
- Hua, G. H., Chen, S. L., Ai, J. T., and Yang, L. G. (2008). None of polymorphism of ovine fecundity major genes *FecB* and *FecX* was tested in goat. *Animal reproduction Sci.* 108 (3–4), 279–286. doi:10.1016/j.anireprosci.2007.08.013
- Jin, M., Lu, J., Fei, X., Lu, Z., Quan, K., Liu, Y., et al. (2020). Selection signatures analysis reveals genes associated with high-altitude adaptation in Tibetan goats from nagqu, tibet. *Anim. (Basel)* 10 (10), 1599. doi:10.3390/ani10091599
- Li, S., Xiang, C., Ge, Y., Liu, H., Zhang, D., and Wang, Z. (2022). Differences in eating quality and electronic sense of meat samples as a function of goat breed and postmortem rigor state. *Food Res. Int. Ott. Ont.* 152, 110923. doi:10.1016/j.foodres.2021.110923
- Liu, Z., Tan, X., Jin, Q., Zhan, W., Liu, G., Cui, X., et al. (2024). Multiomics analyses of Jining Grey goat and Boer goat reveal genomic regions associated with fatty acid and amino acid metabolism and muscle development. *Anim. Biosci.* 37 (6), 982–992. doi:10.5713/ab.23.0316
- McNulty, S. E., Barrett, R. M., Vogel-Ciernia, A., Malvaez, M., Hernandez, N., Davatolhagh, M. F., et al. (2012). Differential roles for Nr4a1 and Nr4a2 in object location vs. object recognition long-term memory. *Learn. and Mem. (Cold Spring Harb. N.Y.)* 19 (12), 588–592. doi:10.1101/lm.026385.112
- Monteagudo, L. V., Arruga, M. V., Bonafonte, J. I., Ordás, M., Whyte, A., Gallego, M., et al. (2008). Bilateral Leydig cell tumor in a six-year-old intersex goat affected by Polled Intersex Syndrome. *Veterinary pathol.* 45 (1), 42–45. doi:10.1354/vp.45-1-42
- Pan, C. Y., Lan, X. Y., Zhao, H. Y., Hu, S. R., Huai, Y. T., Lei, C. Z., et al. (2011). A novel genetic variant of the goat *Six6* gene and its association with production traits in Chinese goat breeds. *Genet. Mol. Res. GMR* 10 (4), 3888–3900. doi:10.4238/2011.November.22.9
- Peng, W., Zhang, Y., Gao, L., Wang, S., Liu, M., Sun, E., et al. (2024). Examination of homozygosity runs and selection signatures in native goat breeds of Henan, China. *BMC genomics* 25 (1), 1184. doi:10.1186/s12864-024-11098-0
- Quan, K., Li, J., Han, H., Wei, H., Zhao, J., Si, H. A., et al. (2020). Review of Huang-huai sheep, a new multiparous mutton sheep breed first identified in China. *Trop. animal health Prod.* 53 (1), 35. doi:10.1007/s11250-020-02453-w
- Simon, R., Lischer, H. E. L., Pieńkowska-Schelling, A., Keller, I., Häfliger, I. M., Letko, A., et al. (2020). New genomic features of the polled intersex syndrome variant in goats unraveled by long-read whole-genome sequencing. *Anim. Genet.* 51 (3), 439–448. doi:10.1111/age.12918
- Wang, J. J., Zhang, T., Chen, Q. M., Zhang, R. Q., Li, L., Cheng, S. F., et al. (2020). Genomic signatures of selection associated with litter size trait in jining gray goat. *Front. Genet.* 11, 286. doi:10.3389/fgene.2020.00286
- Wang, W., Sun, B., Hu, P., Zhou, M., Sun, S., Du, P., et al. (2019). Comparison of differential flavor metabolites in meat of lubei white goat, jining gray goat and boer goat. *Metabolites* 9 (9), 176. doi:10.3390/metabo9090176
- Xiao, J., Liu, B., Yao, Y., Guo, Z., Jia, H., Kong, L., et al. (2022). Wheat genomic study for genetic improvement of traits in China. *Sci. China. Life Sci.* 65 (9), 1718–1775. doi:10.1007/s11427-022-2178-7
- Xiao, Y., Zheng, X., Li, G., Zhou, C., Wu, C., Xu, Z., et al. (2020). Investigation of the effects of dichlorvos poisoning on AMPK signaling pathway in chicken brain tissues. *Environ. Pollut. (Barking, Essex 1987)* 261, 114109. doi:10.1016/j.envpol.2020.114109
- Yang, S., Han, H., Li, J., Zhang, Y., Zhao, J., Wei, H., et al. (2021). Transcriptomic analysis of gene expression in normal goat ovary and intersex goat gonad. *Reproduction Domest. animals* 56 (1), 12–25. doi:10.1111/rda.13844
- Yao, Z., Zhang, S., Wang, X., Guo, Y., Xin, X., Zhang, Z., et al. (2023). Genetic diversity and signatures of selection in BoHuai goat revealed by whole-genome sequencing. *BMC genomics* 24 (1), 116. doi:10.1186/s12864-023-09204-9
- Zhang, W., Wang, L., Hu, B., Jin, M., and Zhou, J. (2024). Changes in ovarian tissue structure and distribution of oestrogen receptors in Huanghuai goats at different ages. *Anat. Histol. Embryol.* 53 (3), e13042. doi:10.1111/ahc.13042
- Zhou, J., Zhang, W., Liu, W., Sheng, J., Li, M., Chen, X., et al. (2020). Histological study of intestinal goblet cells, IgA, and CD3+ lymphocyte distribution in Huang-huai white goat. *Folia Morphol.* 79 (2), 303–310. doi:10.5603/FM.a2019.0082
- Zhou, L., Yang, B., Xue, N., Li, F., Seip, H. M., Cong, X., et al. (2014). Ecological risks and potential sources of heavy metals in agricultural soils from Huanghuai Plain, China. *Environ. Sci. Pollut. Res. Int.* 21 (2), 1360–1369. doi:10.1007/s11356-013-2023-0



OPEN ACCESS

EDITED BY

Fei Hao,
Northumbria University, United Kingdom

REVIEWED BY

Antoine Fages,
Université Toulouse III Paul Sabatier, France
Wei Shi,
Chinese Academy of Sciences (CAS), China
Sarah Meirelles,
Universidade Federal de Lavras, Brazil

*CORRESPONDENCE

Kairat Iskhan,
✉ kayrat_ishan@mail.ru
Gemingguli Muhatai,
✉ gmgl-113@foxmail.com

RECEIVED 27 May 2024

REVISED 26 August 2025

ACCEPTED 26 August 2025

PUBLISHED 12 November 2025

CITATION

Tang C, Yang B, Dawulietihan G, Xue L, Liu S, Yalimaimaiti Y, Wang Q, Yang N, Sun X, Wang Y, Wumaier A, Khizat S, Assanbayev T, Kozhanov Z, Attokurov K, Obdunov E, Li H, Reheman A, Zhou X, Aizimu W, Iskhan K and Muhatai G (2025) The genetic diversity and population structure of native horse breeds in Xinjiang, China.
Front. Genet. 16:1439312.
doi: 10.3389/fgene.2025.1439312

COPYRIGHT

© 2025 Tang, Yang, Dawulietihan, Xue, Liu, Yalimaimaiti, Wang, Yang, Sun, Wang, Wumaier, Khizat, Assanbayev, Kozhanov, Attokurov, Obdunov, Li, Reheman, Zhou, Aizimu, Iskhan and Muhatai. 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 genetic diversity and population structure of native horse breeds in Xinjiang, China

Chi Tang^{1,2}, Baoyu Yang^{1,2}, Gulibaheti Dawulietihan³, Li Xue⁴, Shuyuan Liu⁵, Yina mujiang Yalimaimaiti⁶, Qingzheng Wang⁷, Na Yang^{1,2,7}, Xiaoyuan Sun⁷, Yaru Wang⁷, Ailifeire Wumaier⁷, Serik Khizat⁸, Tolegen Assanbayev⁹, Zhassulan Kozhanov¹⁰, Kursantbek Attokurov¹¹, Elmurat Obdunov¹¹, Hangsen Li⁷, Aikebaier Reheman⁷, Xiaoling Zhou⁷, Wumaierjiang Aizimu^{2,7}, Kairat Iskhan^{8*} and Gemingguli Muhatai^{1,2,7*}

¹College of Life Science and Technology, Tarim University, Alar, Xinjiang, China, ²Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Alar, Xinjiang, China, ³Agricultural (Animal Husbandry) Development Service in Tuerhong Township, Fuyun, Xinjiang, China, ⁴Animal Husbandry Workstation of Fuyun County, Fuyun, Xinjiang, China, ⁵Animal Husbandry Workstation of Balikun County, Balikun, Xinjiang, China, ⁶Animal Husbandry and Veterinary Station of Kalayagazi Town, Yining, Xinjiang, China, ⁷Key Laboratory of Tarim Animal Husbandry Science and Technology, Xinjiang Production & Construction Corps, Alar, Xinjiang, China, ⁸Physiology, Morphology and Biochemistry, Kazakh National Agrarian Research University, Almaty, Kazakhstan, ⁹Zootechnology and Veterinary Medicine, Toraighyrov University, Pavlodar, Kazakhstan, ¹⁰Horse Breeding Department, Kazakh Research Institute of Livestock and Forage Production, Almaty, Kazakhstan, ¹¹Osh State University, Osh, Kyrgyzstan

Introduction: Xinjiang is a region renowned for its rich diversity of native horse breeds, making it one of the most affluent equine genetic resource areas in China. While prized for their high adaptability and tolerance to roughage, the conservation of these native breeds faces challenges from the introduction of external breeds and industrial changes. Furthermore, the unknown population structure of Xinjiang horse breeds has hindered effective conservation efforts.

Methods: This study presents the first comprehensive Single Nucleotide Polymorphism (SNP) analysis of seven Xinjiang native horse breeds. We utilized 10X whole-genome sequencing to assess their genetic diversity, population structure, and genetic relationships.

Results: Our findings revealed a high level of population genetic diversity among the Xinjiang native horse breeds. These breeds exhibited significant genetic differentiation from other horse breeds originating from Europe, Central Asia, Western Asia, and other parts of China. Evidence of frequent historical gene flow was detected, particularly among breeds in northern Xinjiang, which were shown to be more closely related to each other.

Discussion: This study elucidates the distribution patterns, evolutionary characteristics, and substantial genetic diversity of Xinjiang's native horse breeds. The results provide crucial insights into their unique genetic background and population history. These findings offer valuable theoretical

support for establishing core conservation groups of local germplasm, guiding future breeding programs for new cultivars, and further exploration of the characteristics inherent to Xinjiang's native horse genetic resources.

KEYWORDS

native horse breed, whole genome sequencing, genetic diversity, population structure, genetic resources conservation

1 Introduction

The domestic horse (*Equus ferus caballus*) was successfully domesticated approximately 5,500 years ago (Librado et al., 2021). However, recent studies have proposed that the modern domestic horse originated in the western Eurasian steppe around 4,200 years ago (Librado et al., 2021). Subsequently, they spread across Eurasia, giving rise to various breeds adapted to diverse geographic and climatic environments. Horse domestication has enhanced human work efficiency and propelled human civilization (Schubert et al., 2014). According to the Food and Agriculture Organization (FAO), there are over 900 horse breeds globally, with 694 native breeds. Among these, 138 are located in Asia and 371 in Europe (<https://www.fao.org/>).

The primary purposes for horse keeping include dairy and meat production, racing, equestrian activities, and maintaining free-range populations. However, agricultural advancements have decreased the overall use of horses, significantly reducing the global horse population. Over the past 2,000 years, domestic horses have lost nearly half of their genetic diversity, with genomic heterozygosity decreasing by approximately 16% in the last 200 years (Fages et al., 2019). Genetic diversity and inbreeding are often connected (Spielman et al., 2004), which can inform the efficacy of breeding conservation and offer insights into the sustainable development and utilization of horse populations. This dramatic decline in the global horse population has severely affected the genetic diversity and number of purebred native breeds. The decreased genetic diversity of domestic horses is primarily due to purebred selection practices, which increase deleterious mutations in domestic horse genomes (Orlando, 2020).

China is known for having a diverse and abundant population of horses, with a rich genetic variety (Ling et al., 2011). Furthermore, Chinese horse breeds include the Northern China (NC), Qinghai-Tibetan (QT), and Southwestern (SW) populations (Liu et al., 2019). The Xinjiang Uygur Autonomous Region (XUAR), a prominent horse-producing area in China, has the largest horse population in the country, reaching 1,107,000 horses as of 2022. This region is known for four native (Kazakh, Yanqi, Kyrgyz, Balikun) and two cultivated (Yili and Yiwu) horse breeds, mainly bred for meat. China is the largest global horse meat producer (FAO statistics, 2022), producing 159,068 tons of horse meat, most of which is produced in Xinjiang. Many horses are underutilized and freely grazed for breeding purposes, with only a small number used for meat or riding (including leisure, competitions, celebrations, etc.). However, the demand for sports events and leisure tourism has changed the use of native horse breeds, with some regions crossbreeding native horses with Thoroughbred and Arabian breeds to achieve better competition results. This practice enhances the genetic diversity and productivity of cultivated horse breeds but significantly threatens the genetic purity of native breeds. Thus, there is a risk of extinction for

native horse breeds with small population sizes due to unorganized selection and the introduction of hybrids.

Sequence analysis of mtDNA from eastern Chinese and European horse populations revealed a higher frequency of haplogroup F and a lower frequency of haplogroup D in eastern Chinese populations (McGahern et al., 2006). In contrast, European populations exhibited the opposite trend, suggesting a form of genetic isolation and differentiation between these two populations (McGahern et al., 2006). This finding was further supported by the mtDNA data from ancient horses in China dating back to 2,000–4,000 years ago (Cai et al., 2009).

Xinjiang, situated in Central Asia, the transitional region between eastern China and Europe, The ancient stone culture in Ili River basin and Boltala River basin in Xinjiang may be related to the early domestication activities (Annie and Dexin, 2020). Xinjiang Kazakh, Yanqi, and Barkun horses exhibit higher Y chromosome diversity than other horse breeds in China, possibly due to the expansive grasslands, traditional animal husbandry practices, and minimal human interference in Xinjiang (Han et al., 2015). Research on the genetic diversity and structure of Xinjiang horse breeds is limited, with most studies focusing on microsatellite markers. However, Whole Genome Sequencing (WGS) has successfully revealed the genetic diversity and structure of various horse breeds, such as the Thoroughbred (Tozaki et al., 2021), Mongolian (Huang et al., 2014), Lichuan, Kazakh (Zhang et al., 2018), and Mavari horses (Jun et al., 2014). Nonetheless, no comprehensive genome-wide study has been used to analyze the genetic information of horse breeds in Xinjiang. There is no molecular data for all the horse breeds in this region.

Therefore, this study investigated the population genetic structure of native horse breeds in Xinjiang. The study utilized WGS technology to gather genetic data from seven native horse breeds (taxa) in Xinjiang, China. The native horse breeds included Kazakh (Altay and Yili region), Yanqi, Kyrgyz, Balikun, and other unexplored breeds, such as the Kunlun (Hotan region, Kunlun Mountains) and Tashkurgan (Tashkurgan Tajik Autonomous County). These horse breeds are divided by the Tianshan Mountains, North of the Tianshan Mountains are the northern horses, which include the Kazakh and Balikun breeds, while the remaining horse breeds are distributed south of the Tianshan Mountains as southern horses. Genetic variations within and between these breeds were analyzed and compared with data from the horse breed sequences in the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>). The study provides the first collection of genetic information from all Xinjiang native horse breeds. These findings will enhance our understanding of the population genetics of native horse breeds and shed light on the history of horse domestication in Xinjiang.

2 Materials and methods

2.1 Sample collection

In this study, 70 peripheral blood samples were collected from seven horse breeds (taxa), with samples taken from ten horses of each breed. The breeds included Kazakh horse in the Altay region (KA), Kazakh horse in the Yili region (KY), Yanqi (YQ), Balikun (BLK), Kyrgyz (KE), Tashkurgan (TX), and Kunlun (KL) horse. To ensure that the samples were representative of the genetic variation of the breeds, all the samples were collected from different remote pastures, with at least two separate populations sampled for each breed, and we used information provided by the owners to avoid the collection of samples from related horses and exclude parents or progeny. Herders were consulted for background information on the selection and breeding history of their native breeds. Subsequently, the genealogical records were reviewed to ensure that the samples were unrelated but representative of the genetic variation of the breeds. Blood samples from 70 domestic horses were used for Genomic DNA extraction using the TaKaRa MiniBEST Whole Blood Genomic DNA Extraction Kit (Takara Bio, Beijing, China).

The genomic DNA concentration was determined using an INVITROGEN Qubit 4.0 Fluorometer (Invitrogen Inc, CA, United States), and its integrity was assessed through agarose gel electrophoresis. Subsequently, the samples were subjected to whole-genome resequencing on the Illumina HiSeqX Ten platform (Illumina Inc, CA, United States) based on the quality of the whole-genome DNA extraction (the nucleic acid concentrations were all greater than 80 ng/μL, with OD260/280 ratios ranging from 1.8 to 2.0). The genomic DNA samples were stored at −80 °C in a freezer. In addition, sequencing data of 32 horses (including nine domestic horse breeds and one wild horse population) were downloaded from the NCBI (Supplementary Table S1). The ten horse breeds included the Akhal-Teke (n = 1), Arabian (n = 2), Curly (n = 2), Debaos pony (n = 5), Mongolian (n = 4), Sorraia (n = 1), Thoroughbred (n = 4), Tibetan (n = 5), Yakut (n = 6), and Przewalski (n = 2).

2.2 Detection and quality control of the SNP sites

The whole genome sequences of 70 domestic horse samples yielded approximately 1831.95 GB of raw data (average of 26 GB/sample), with the Q20 and Q30 base quality exceeding 95.00% and 90.00%, respectively (Supplementary Table S1). After removing low-quality reads, each horse genome had a 10-fold coverage. The high-quality sequencing data from 70 horses, along with the sequencing data of 32 horses obtained from NCBI, were mapped to the domestic horse reference genome, EquCab3.0 (Kalbfleisch et al., 2018). Using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). Eight samples with low matching rates were removed (Supplementary Table S2). Next, SAMtools v1.3 (Li et al., 2009) and GATK v3.1 (DePristo et al., 2011) were utilized to detect SNPs at variant sites. The 62 equine SNP datasets were merged with the 32 SNP datasets from the NCBI using PLINK v1.9 (Chang et al., 2015). Subsequently, SNP loci with <90% detection rates, a <0.05 minor allele frequency (MAF), and >1% missing genotype rates were filtered out using pi-

hat in PLINK v1.9 (Chang et al., 2015), and the pedigree relationships between samples were evaluated (Petersen et al., 2013). The filter parameters were as follows: (1) Fisher test of strand bias (FS) ≤ 60; (2) HaplotypeScore ≤ 13.0; (3) Mapping Quality (MQ) ≥ 40; (4) Quality Depth (QD) ≥ 2; (5) ReadPosRankSum ≥ −8.0; and (6) MQRankSum > −12.5 (Supplementary Table S3; Supplementary Figures S1–S3). Linkage disequilibrium (LD) was evaluated by calculating r^2 at the genome-wide intervals of 25 SNPs with a window size of 100 SNPs. The dataset was further filtered to remove samples with low quantities, retaining others representative of the five horse breeds, using a 0.2 threshold for judging LD (--indep-pairwise 100 25 0.2).

2.3 Analysis of genetic diversity and population structure

The expected heterozygosity (He) and coefficient of inbreeding (Fis) were calculated using the PLINK software. The He for each population was performed through the Hardy-Weinberg test, while Fis was calculated per sample using the number of pure heterozygotes. Furthermore, the principal component analysis (PCA) and population differentiation index (Fst) were determined using smartPCA in the EIGENSOFT software v4.2 (Price et al., 2006). The proportion of missing data to be ≤10% was allowed. Phylogenetic trees were constructed using the neighbor-joining (NJ) method in FastTree v2.1.11 (Price et al., 2009). The resulting tree file was subsequently uploaded to the iTol v5 (Letunic and Bork, 2021) for visualization purposes. Phylogenetic networks were visualized in R version 3.5.1. The genetic structure of the population was assessed with the fastSTRUCTURE package (Raj et al., 2014). The results of the fastSTRUCTURE analysis were then visualized using the DISTRUCT software (Rosenberg et al., 2002).

3 Results

The four main Xinjiang horse breeds, developed in diverse environmental conditions, exhibited distinct characteristics (Figure 1). The 94 samples that met the specified quality control criteria generated 26,539,717 SNPs (Supplementary Table S4).

3.1 Analysis of genetic diversity

Table 1 shows that the heterozygosity values across all the breeds and populations were notably similar. The observed heterozygosity (Obs Het) reflects the actual proportion of heterozygosity within the population. The genetic diversity of the Xinjiang native horse breeds was found to be relatively low compared to Mongolian and Tibetan horse (Supplementary Table S5). Furthermore, the Obs Het for all the populations was lower than the expected heterozygosity (Exp Het), suggesting the presence of varying degrees of inbreeding or genetic drift. Additionally, Fisher's inbreeding coefficient (Fis) provided further insights into the degree of inbreeding; notably, the YQ population exhibited the highest Fis value (0.1844). The Fis results for the Xinjiang native horse breeds indicated generally low levels of inbreeding, with the lowest coefficients observed in the KE

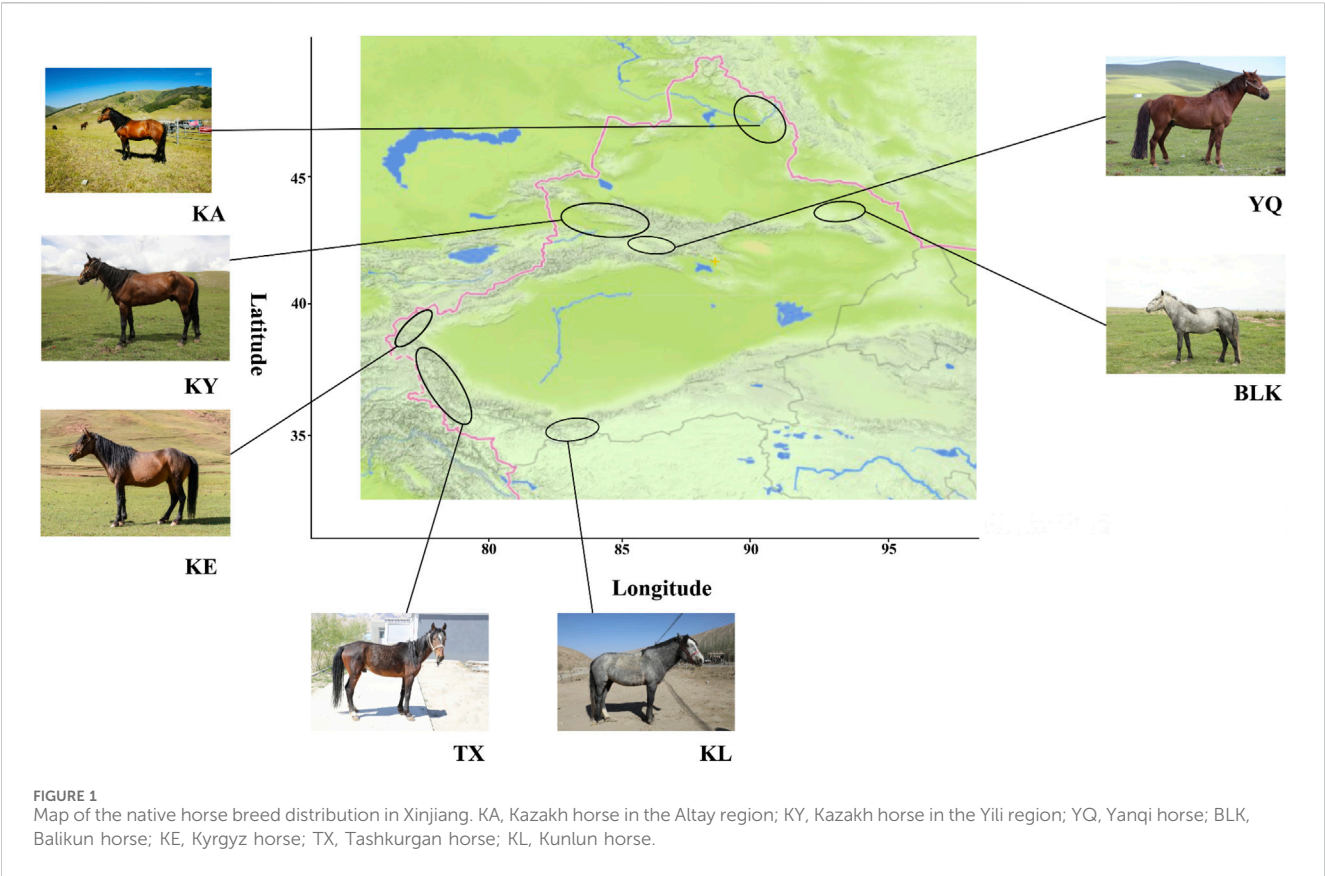
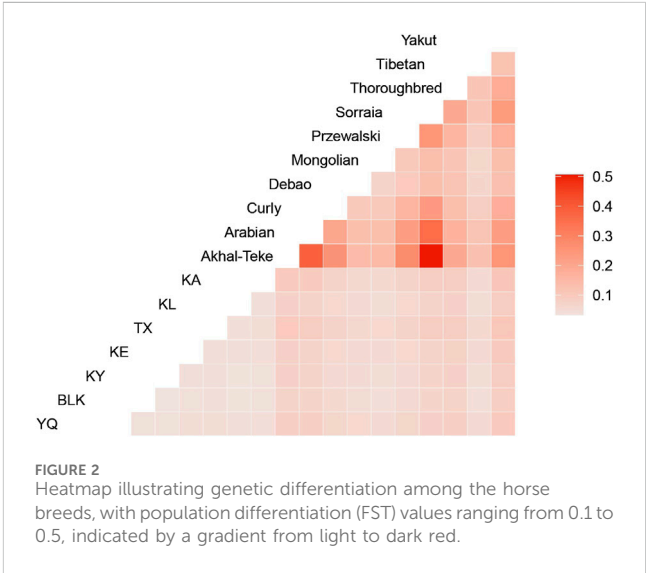


TABLE 1 Comparison of genetic diversity of Xinjiang horse breeds.

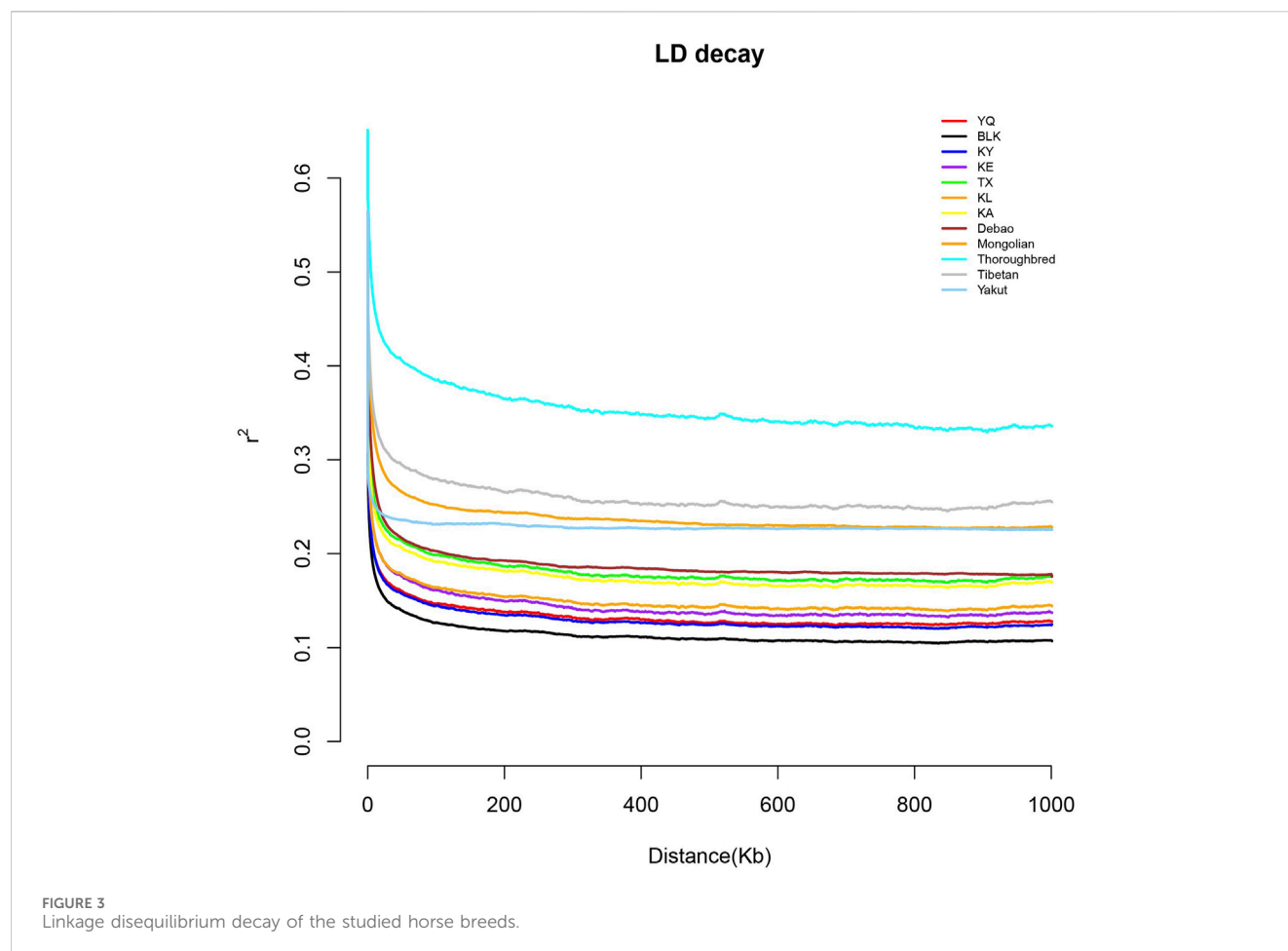
Breed/Population	Obs Het	Exp Het	Pi	Fis
YQ	0.1940	0.2481	0.2632	0.1844
BLK	0.2207	0.2541	0.2682	0.1336
KY	0.2330	0.2543	0.2697	0.1013
KE	0.2294	0.2486	0.2619	0.0915
TX	0.2248	0.2472	0.2668	0.1061
KL	0.2408	0.2536	0.2689	0.0781
KA	0.2285	0.2490	0.2693	0.1026

and KL populations, likely attributed to differences in stallion distributions within these groups. The polymorphic information content (PIC/Pi) value of 0.2689 indicates that the KL population exhibits higher levels of polymorphism and more pronounced genetic differences among individuals. This elevated Pi value reflects a greater genetic diversity, which is crucial for the adaptation of the assessment population. Understanding this genetic diversity is significant for breeding strategies and overall population management.

The genetic differentiation index (F_{st}) between the horse groups demonstrates the genetic divergence among the different horse populations. The differentiation within the Xinjiang native horse breeds was relatively lower than that of the other breeds. The F_{st} values for the Xinjiang horse breeds did not surpass 0.01, indicating



minimal differentiation among the samples. In contrast, the Arabian and Akhal-Teke horses exhibited higher differentiation from the remaining groups, with the highest F_{st} compared with the other horse breeds (Figure 2). Although Przewalski's horses diverged from domestic horses approximately 40,000 years ago, their F_{st} values with domestic breeds were lower than expected (Figure 2). This pattern may reflect shared ancestral polymorphisms at conserved neutral loci, as captured by the SNP array. Additionally,



demographic histories, such as bottlenecks during domestication, may have contributed to the homogenization of diversity at certain loci. This may be related to their breeding strategies. Both horse breeds have undergone long-term closed breeding and intensive selection to preserve their unique and exceptional traits, such as the endurance and elegance of Arabian horses, as well as the speed and endurance of Akhal-Teke horses. While these practices maintain the desired breeding characteristics, they also result in a decrease in genetic diversity and an increase in interpopulation differentiation.

The Xinjiang native horse breeds had consistently lower LD Decay rates than the domestic native and foreign breeds (Figure 3), suggesting a higher genetic diversity than non-Chinese breeds retrieved from NCBI, with BLK having the highest value in this study. In contrast, Thoroughbreds, having undergone extensive and systematic selective breeding and controlled reproduction to meet specific performance requirements, especially for racing, had the lowest genetic diversity.

3.2 Genetic structure

The PCA revealed limited differentiation among the Xinjiang native horse breeds. These native breeds clustered together with considerable overlap (Figure 4). The results of the Principal Component Analysis (PCA) presented in Figure 4A demonstrate

a distinct separation between the native horse breeds of China and the foreign horse breeds, utilizing Przewalski's horse as the reference population. Among them, the Yakut, Thoroughbred, Sorraia, and Curly horse breeds were obviously different from Xinjiang native horse breeds, while Thoroughbred and Yakut showing notable separation from the other horse breeds. Figure 4B shows the distribution of these Xinjiang native horse breeds on principal component PC1 and PC2, with different colored ellipses indicating the 95% confidence intervals for each population. The size and shape of these ellipses reflect the genetic differences between the populations. Among them, the confidence ellipses of KA and KE were smaller and overlapping, indicating that they had a similar genetic background. The Xinjiang native horse breeds are closest in genetic background to Mongolian and Tibetan horses, possibly due to the historical exchanges between these breeds. Furthermore, the PCA results revealed significant genetic differentiation between the Central and Western Asian breeds, including Arabian and Akhal-Teke horses, corroborating the F_{st} results.

The native horses of Xinjiang were classified into a single large taxon, indicating that these horses are largely distinct from non-Xinjiang breeds. This pattern is more likely due to the long-term independent evolutionary history of Xinjiang horses, which has led to the accumulation of unique mutations over generations, rather than solely due to unique genetic characteristics during early domestication. (Figure 5A). Moreover, the Debao horse, which

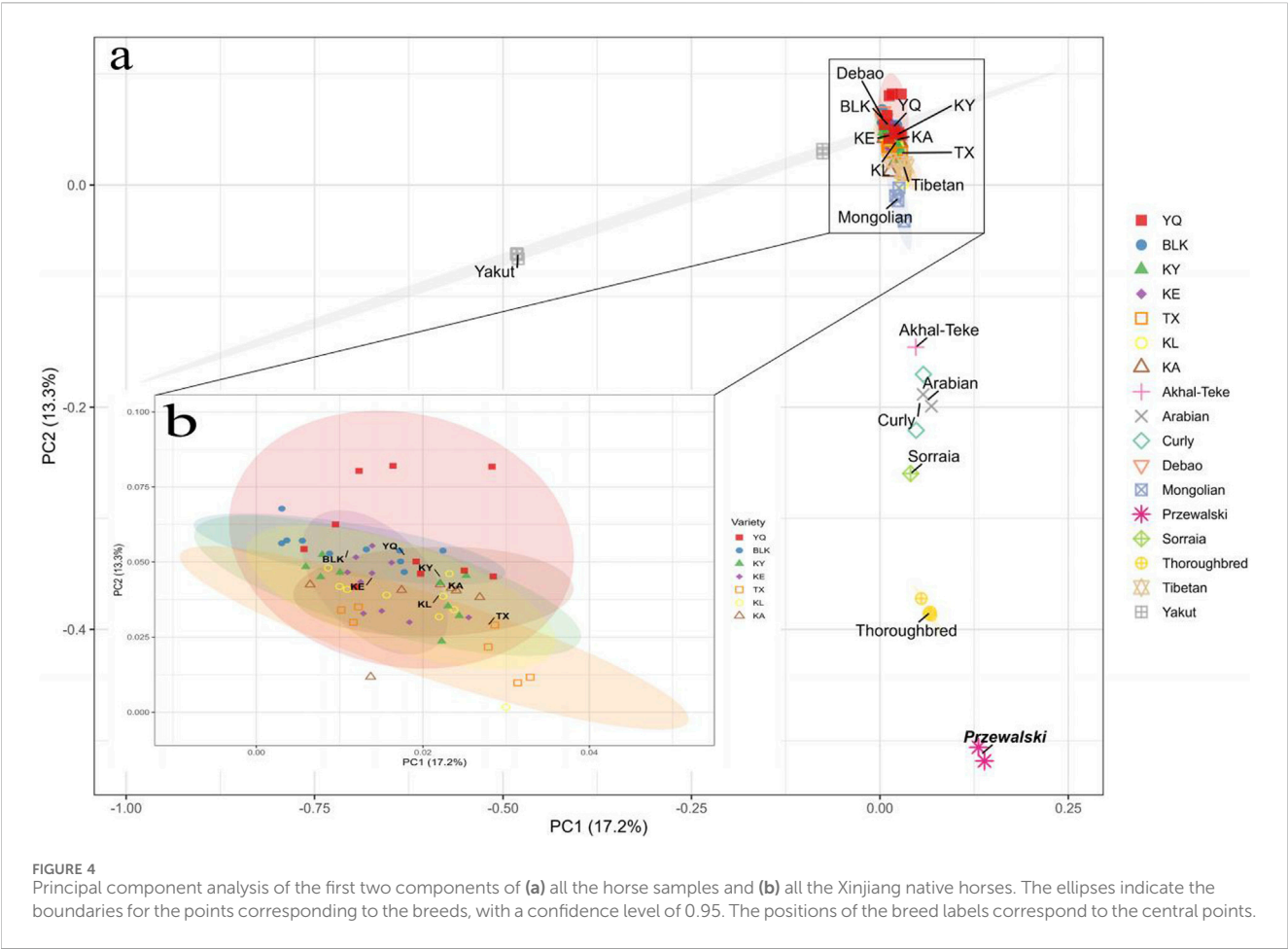


FIGURE 4 Principal component analysis of the first two components of (a) all the horse samples and (b) all the Xinjiang native horses. The ellipses indicate the boundaries for the points corresponding to the breeds, with a confidence level of 0.95. The positions of the breed labels correspond to the central points.

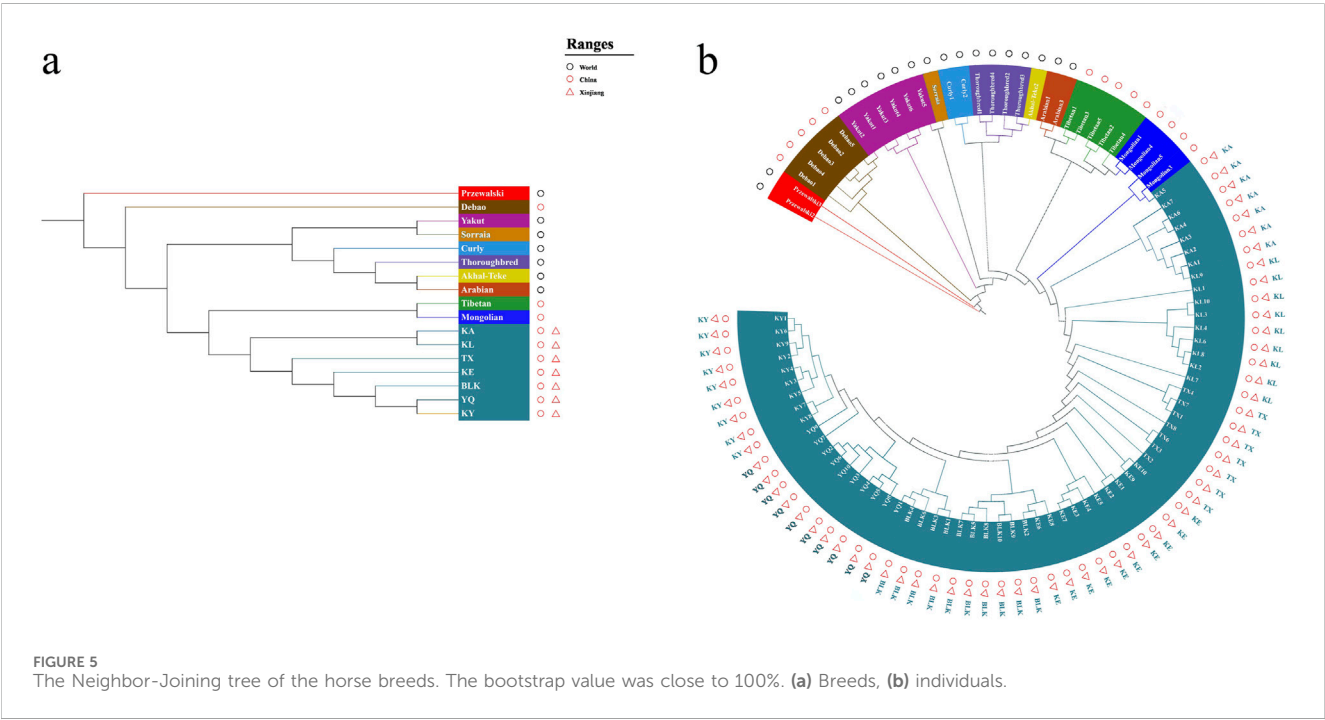


FIGURE 5 The Neighbor-Joining tree of the horse breeds. The bootstrap value was close to 100%. (a) Breeds, (b) individuals.

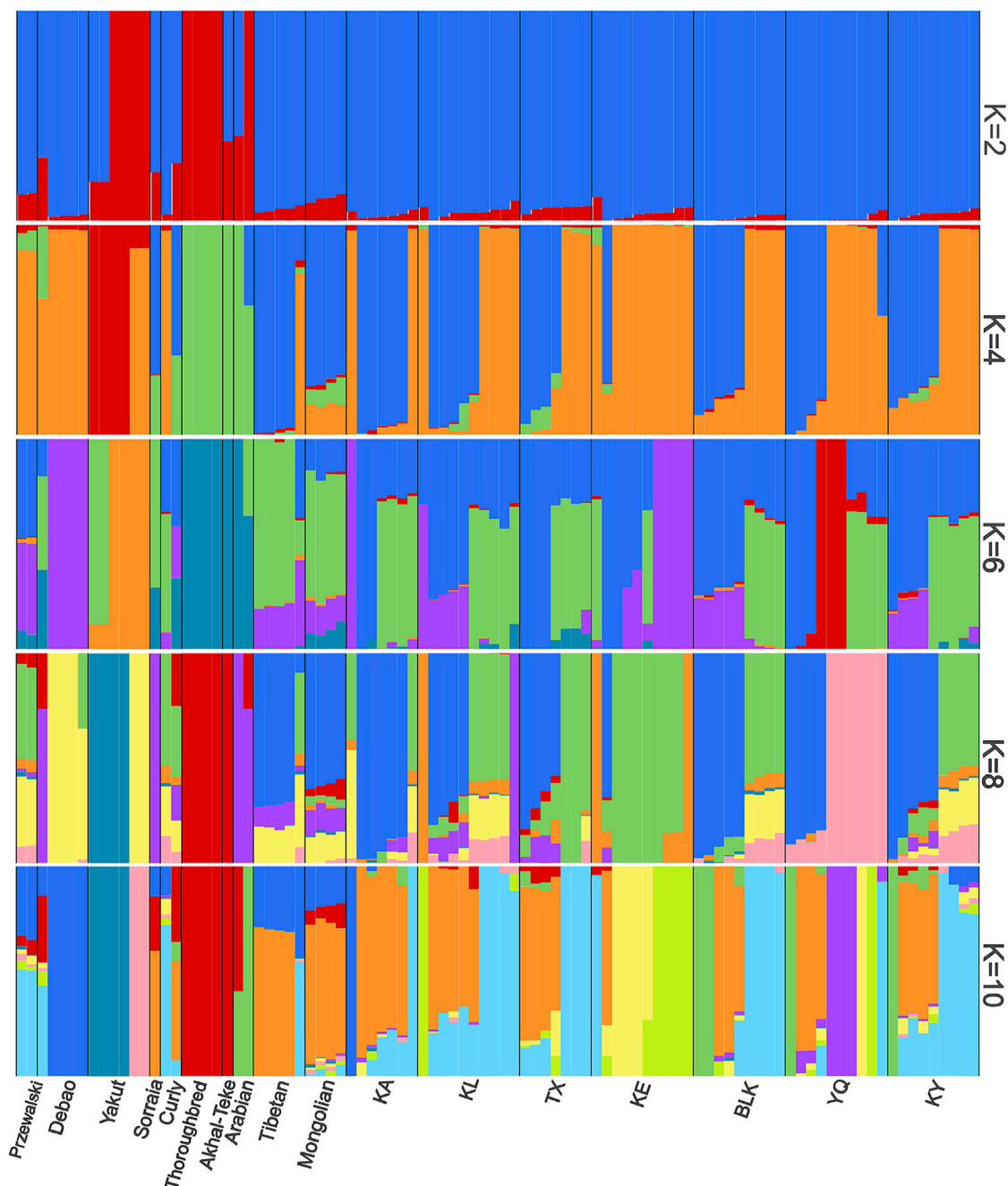


FIGURE 6

Comparative structure of the horse populations with the number of ancestral clusters (K) ranging from 2 to 10. Each bar represents an individual for each breed.

exhibits a unique evolutionary pattern characterized by a distinct genetic lineage, stands out from other breeds. Meanwhile, the remaining Chinese and foreign horse breeds can be broadly categorized into two major genetic clusters. Subsequent classification identified two main groups in southern and northern Xinjiang, highlighting a significant genetic distance between the breeds in these regions and reflecting the geographic distance (Figure 5B). The Debao pony exhibited the most primitive branching, while the Mongolian and Tibetan horses displayed a closer phylogenetic relationship. In Xinjiang, KA, KY, YQ, and BLK

were closely related to the northern Xinjiang breeds, whereas KE and TX were more closely related to the southern Xinjiang breeds, which is consistent with the findings of the PCA results.

The population structure was analyzed for all the horse samples, considering a K value between 2 and 10, representing the ancestral groups (Figure 6). At a value of $K = 4$, we could distinguish the separate groups of horse breeds in China and horse breeds in other countries (Supplementary Figure S4). The population genetic structure analysis revealed that at $K = 6$, the Xinjiang native horse breeds were distinct. The analysis of $K = 10$ reveals that

the genetic structure of the native horse breed in Xinjiang is highly complex. This complexity arises not only from random genetic drift or mutations but is more likely attributable to prolonged gene flow.

4 Discussion

This study included seven Xinjiang native (KY, YQ, BLK, KE, TX, and KL), three Chinese (Mongolian, Tibetan, and Debao pony), and six foreign (Thoroughbred, Curly, Yakut, Sorraia, Arabian, and Akhal-Teke) horse breeds. Xinjiang native horse breeds exhibited typical genetic characteristics that low inbreeding levels, aligning with the traditional open breeding practices on Xinjiang grasslands. This result suggests strong resistance to inbreeding within this population. Furthermore, Thoroughbreds had the lowest genetic diversity due to varying degrees of bottleneck effects caused by closed artificial selection and high levels of inbreeding (Khanshour et al., 2013; Cosgrove et al., 2020). Ethnic groups in Xinjiang prefer specific horse coat colors; for example, Kazakhs favor dark colors (red bay and black bay), and Tajiks favor light color horses. Therefore, selections are based on coat color preferences. Hence, Thoroughbred and Arabian crosses with desired coat colors were introduced. Interestingly, the genetic structure, which was influenced by coat color preferences, has maintained the genetic diversity of the horse breeds (Druml et al., 2009).

Xinjiang horse breeds are genetically distinct from other Chinese, Central Asian, and European horse breeds. This differentiation is likely primarily driven by the long-term independent evolutionary history following the divergence of these horse breeds from other populations. Additionally, other factors such as historical population bottleneck effects, genetic drift, special environmental adaptability, and targeted artificial selection for specific traits (Castaneda et al., 2019) may have further contributed to their genetic distinctiveness. Moreover, Xinjiang horse breeds have a relatively independent genetic structure and are more closely related to Tibetan and Mongolian horses. The genetic makeup of Xinjiang horse breeds, particularly Kazakh horses, aligns with previous research findings (Liu et al., 2019). Crossbreeding practices in Xinjiang involve using native horse breeds as dams and Arabian or Thoroughbred horses as sires to produce offspring with enhanced speed and endurance for competitive events. Thus, this study included all purebred native breeds, suggesting that other breeds were introduced in the past for crossbreeding to improve performance, albeit with challenges in tracing their pedigree and genealogical information due to insufficient records. This study, based on a genetic analysis of 32 horses from outside Xinjiang using genome-wide autosomal SNPs, revealed significant genetic distinctions, reflecting their geographic origin and breed history (Petersen et al., 2013). Conversely, studies on the Kazakh horse mtDNA indicated that native horse breeds in western China have multiple matrilineal origins and exchange genetic material with each other and other horse breeds during mobility (Zhang et al., 2012). The PCA and structural results from this study indicate that Kazakhstan possesses a complex genetic background, with both autosomal and mitochondrial DNA (Gemingguli et al., 2016) findings corroborating each other.

Historical interbreeding probably influenced horse breeds divergence. Archaeological data revealed significant gene flow between populations in the Pamir Plateau and Ferghana Valley, acting as a 'transmitter' for exchanges between eastern (Tarim Basin, China) and western (Bactria, Uzbekistan) populations (Hemphill and Mallory, 2004). Horse breeds in the Middle East are closely related to local horse breeds in Xinjiang and may be influenced by human activities. Future research should investigate the gene flow between these populations in conjunction with historical events.

In our analysis, we observed distinct genetic clusters among different horse breeds, which highlights the genetic differentiation between breeds. However, it is important to note that within each breed, there may exist multiple distinct subgroups with significant genetic differences. This intrabreed heterogeneity could be due to factors such as regional differences, selective breeding practices, or historical gene flow events. Genetic differentiation has been observed within various native horse breeds in Xinjiang, particularly in the Northern Xinjiang region. This differentiation is primarily attributed to the introduction of foreign germplasm for crossbreeding. Since herders often introduced Thoroughbreds for crossbreeding, this could seriously affect the purebred germplasm resources and reduce the genetic diversity of the native horse population in Xinjiang.

5 Conclusions

This study presents the first comprehensive analysis of whole genome sequences of Xinjiang native horse breeds, revealing the genetic diversity and structure of these breeds. Xinjiang native horse breeds exhibited higher genetic diversity, with evidence of gene exchange and similarities in the genetic backgrounds of the different groups. The genetic structure of the southern and northern border breeds differed. Additionally, Tashkurgan and Kunlun horses may represent new cryptic horse breeds. Phylogenetic analysis shows that Tashkurgan and Kunlun horses form a distinct clade separate from other Xinjiang breeds (see Figure 5). These horses also exhibit unique genetic signatures in their SNP profiles, indicating a degree of genetic isolation and differentiation. Further investigation is warranted to fully characterize these potential cryptic breeds. Our findings assessed genetic variability and inbreeding within these horse breeds, thereby laying a foundation for future horse breeding policies. They could inform the delineation of protected areas and breeding grounds to prevent large-scale inbreeding and hybridization and will provide a valuable reference for the conservation and utilization of native horse breeds in Xinjiang.

Data availability statement

The data presented in the study are deposited in the NCBI SRA repository, accession number PRJNA1185891.

Ethics statement

The animal studies were approved by Institutional Animal Care and Use Committee at Tarim University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

CT: Writing – original draft, Writing – review and editing. BY: Writing – original draft. GD: Writing – original draft. LX: Resources, Supervision, Writing – review and editing. SL: Resources, Supervision, Writing – review and editing. YY: Resources, Supervision, Writing – review and editing. QW: Data curation, Investigation, Writing – review and editing. NY: Data curation, Investigation, Writing – review and editing. XS: Formal Analysis, Writing – review and editing. YW: Formal Analysis, Writing – review and editing. AW: Resources, Writing – review and editing. SK: Methodology, Writing – review and editing. TA: Methodology, Writing – review and editing. ZK: Methodology, Writing – review and editing. KA: Methodology, Writing – review and editing. EO: Methodology, Writing – review and editing. HL: Resources, Writing – review and editing. AR: Formal Analysis, Writing – review and editing. XZ: Conceptualization, Writing – review and editing. WA: Methodology, Writing – review and editing. KI: Conceptualization, Project administration, Supervision, Writing – review and editing. GM: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China [31960643, 31560616]; Graduate Research Innovation Project of Tarim University [TDBSCX202210]; Open project of Xinjiang

Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin [BRZD2203]; Agricultural germplasm resources survey, collection, protection and identification service project of the Ministry of Agriculture and Rural Affairs [19221122].

Acknowledgments

The authors gratefully thanks to the Agriculture and Rural Bureau of Fuyun, Nileke, Barikun, Hotan, Tashkurgan, Wuqia, Hejing County, etc. for their support. The computations in this paper were run on the bioinformatics computing platform of Tarim University. We are also grateful to Prof. Indira Beishova (Zhangir Khan West-Kazakhstan Agrarian-Technical University), Prof. Malika Shamekova (Institute of Plant Biology and Biotechnology), and Prof. Abdugani Abdurasulov (Osh State University) for their constructive comments on the 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.2025.1439312/full#supplementary-material>

References

- Annie, C., and Dexin, C. (2020). Results of field research on ancient stonework in the river valleys of bortala and ili in Western tian Shan (xinjiang, China). *Asian Perspect.* 59, 385–420. doi:10.1353/asi.2020.0019
- Cai, D., Tang, Z., Han, L., Speller, C. F., Yang, D. Y., Ma, X., et al. (2009). Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J. Archaeol. Sci.* 36, 835–842. doi:10.1016/j.jas.2008.11.006
- Castaneda, C., Juras, R., Khanshour, A., Randalht, I., Wallner, B., Rigler, D., et al. (2019). Population genetic analysis of the Estonian native horse suggests diverse and distinct genetics, ancient origin and contribution from unique patrines. *Genes (Basel)* 10, 629. doi:10.3390/genes10080629
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8
- Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., et al. (2020). genome diversity and the origin of the arabian horse. *Sci. Rep.* 10, 9702. doi:10.1038/s41598-020-66232-1
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi:10.1038/ng.806
- Druml, T., Baumung, R., and Sölkner, J. (2009). Pedigree analysis in the Austrian noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. *J. Anim. Breed. Genet.* 126, 348–356. doi:10.1111/j.1439-0388.2008.00790.x
- Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., et al. (2019). tracking five millennia of horse management with extensive ancient genome time series. *Cell* 177, 1419–35.e31. doi:10.1016/j.cell.2019.03.049
- Gemingguli, M., Iskhan, K. R., Li, Y., Qi, A., Wunirifu, W., Ding, L. Y., et al. (2016). Genetic diversity and population structure of Kazakh horses (*Equus caballus*) inferred from mtDNA sequences. *Genet. Mol. Res.* 15. doi:10.4238/gmr.15048618
- Han, H., Zhang, Q., Gao, K., Yue, X., Zhang, T., Dang, R., et al. (2015). Y-Single nucleotide polymorphisms diversity in Chinese Indigenous horse. *Asian-Australas J. Anim. Sci.* 28, 1066–1074. doi:10.5713/ajas.14.0784

- Hemphill, B. E., and Mallory, J. P. (2004). Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from Western central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am. J. Phys. Anthropol.* 124, 199–222. doi:10.1002/ajpa.10354
- Huang, J., Zhao, Y., Shiraigol, W., Li, B., Bai, D., Ye, W., et al. (2014). Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci. Rep.* 4, 4958. doi:10.1038/srep04958
- Jun, J., Cho, Y. S., Hu, H., Kim, H. M., Jho, S., Gadhvi, P., et al. (2014). Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics* 15 (Suppl. 9), S4. doi:10.1186/1471-2164-15-S9-S4
- Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Jr., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., et al. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* 1, 197. doi:10.1038/s42003-018-0199-z
- Khanshour, A., Conant, E., Juras, R., and Cothran, E. G. (2013). Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *J. Hered.* 104, 386–398. doi:10.1093/jhered/est003
- Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi:10.1093/nar/gkab301
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698
- 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
- Librado, P., Khan, N., Fages, A., Kusliy, M. A., Suchan, T., Tonasso-Calvière, L., et al. (2021). The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 598, 634–640. doi:10.1038/s41586-021-04018-9
- Ling, Y. H., Ma, Y. H., Guan, W. J., Cheng, Y. J., Wang, Y. P., Han, J. L., et al. (2011). Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Anim. Genet.* 42, 56–65. doi:10.1111/j.1365-2052.2010.02067.x
- Liu, X., Zhang, Y., Li, Y., Pan, J., Wang, D., Chen, W., et al. (2019). EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol. Biol. Evol.* 36, 2591–2603. doi:10.1093/molbev/msz158
- Mcgahern, A., Bower, M. A. M., Edwards, C. J., Brophy, P. O., Sulimova, G., Zakharov, I., et al. (2006). Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Anim. Genet.* 37, 494–497. doi:10.1111/j.1365-2052.2006.01495.x
- Orlando, L. (2020). The evolutionary and historical foundation of the modern horse: lessons from Ancient genomics. *Annu. Rev. Genet.* 54, 563–581. doi:10.1146/annurev-genet-021920-011805
- Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., et al. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* 8, e54997. doi:10.1371/journal.pone.0054997
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi:10.1038/ng1847
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi:10.1093/molbev/msp077
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi:10.1534/genetics.114.164350
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science* 298, 2381–2385. doi:10.1126/science.1078311
- Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., et al. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5661–E5669. doi:10.1073/pnas.1416991111
- Spielman, D., Brook, B. W., and Richard, F. (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15261–15264. doi:10.1073/pnas.0403809101
- Tozaki, T., Ohnuma, A., Kikuchi, M., Ishige, T., Kakoi, H., Hirota, K. I., et al. (2021). Rare and common variant discovery by whole-genome sequencing of 101 Thoroughbred racehorses. *Sci. Rep.* 11, 16057. doi:10.1038/s41598-021-95669-1
- Zhang, T., Lu, H., Chen, C., Jiang, H., and Wu, S. (2012). Genetic diversity of mtDNA D-loop and maternal origin of three Chinese native horse breeds. *Asian-Australas J. Anim. Sci.* 25, 921–926. doi:10.5713/ajas.2011.11483
- Zhang, C., Ni, P., Ahmad, H. I., Geringguli, M., Baizilaitibei, A., Gulibaheti, D., et al. (2018). Detecting the population structure and scanning for signatures of selection in horses (*Equus caballus*) from whole-genome sequencing data. *Evol. Bioinform Online* 14, 1176934318775106. doi:10.1177/1176934318775106

Frontiers in Veterinary Science

Transforms how we investigate and improve
animal health

The third most-cited veterinary science journal,
bridging animal and human health with a
comparative approach to medical challenges. It
explores innovative biotechnology and therapy for
improved health outcomes.

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

