frontiers | Research Topics

Epigenomics implication for economic traits in domestic animals

Edited by
Chunfang Zhao, Zebin Zhang, Bo Han, Guoqiang Yi, Shenghe Li and Sihua Jin

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-8325-3123-5
DOI 10.3389/978-2-8325-3123-5

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

Epigenomics implication for economic traits in domestic animals

Topic editors

Chunfang Zhao — Anhui Science and Technology University, China
Zebin Zhang — Uppsala University, Sweden
Bo Han — China Agricultural University, China
Guoqiang Yi — Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China
Shenghe Li — Anhui Science and Technology University, China
Sihua Jin — Anhui Agricultural University, China

Citation

Zhao, C., Zhang, Z., Han, B., Yi, G., Li, S., Jin, S., eds. (2023). *Epigenomics implication for economic traits in domestic animals*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3123-5

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 17 July 2023
DOI 10.3389/fgene.2023.1252640

OPEN ACCESS

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

*CORRESPONDENCE
Chunfang Zhao,
zhyy199002@163.com

RECEIVED 04 July 2023
ACCEPTED 11 July 2023
PUBLISHED 17 July 2023

CITATION
Zhao C, Han B, Yi G, Jin S, Zhang Z and Li S (2023), Editorial: Epigenomics implication for economic traits in domestic animals. *Front. Genet.* 14:1252640. doi: 10.3389/fgene.2023.1252640

Editorial: Epigenomics implication for economic traits in domestic animals

Chunfang Zhao[1,2]*, Bo Han[3], Guoqiang Yi[4], Sihua Jin[5], Zebin Zhang[6] and Shenghe Li[1,2]

[1]College of Animal Science, Anhui Science and Technology University, Chuzhou, Anhui, China, [2]Anhui Province Key Laboratory of Animal Nutritional Regulation and Health, Chuzhou, Anhui, China, [3]Department of Animal Genetics and Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, [4]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, [5]College of Animal Science and Technology, Anhui Agricultural University, Hefei, China, [6]Division of Plant Ecology and Evolution, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

KEYWORDS

domestic animals, epigenomics modification, economic traits, ancient DNA, genetic markers

Editorial on the Research Topic
Epigenomics implication for economic traits in domestic animals

Disease resistance, reproductive performance, lactation performance, meat production, egg production, down production, meat quality, egg quality, and other traits are crucial for the economic performance of the livestock farming industry and the quality of livestock products. Epigenetic modifications have a significant impact on the expression of these traits, including DNA methylation, histone modifications, chromatin remodelling, and non-coding RNAs that regulate gene expression. Epigenetics is involved in various biological processes, including embryonic development, tissue differentiation, and tumourigenesis. Environmental stress can influence epigenetic modifications, which are stored in the genome through chromatin modifications and high-level structures. Factors such as nutrition, feeding practices, and environmental factors can influence epigenetic modifications. As biotechnology advances, the development of experimental and analytical methods in epigenetic modifications becomes increasingly important.

In this Research Topic, we cover genetic markers associated with the economic traits of different livestock breeds, various epigenetic modifications that affect livestock economic traits, and the applications of different epigenetic modifications in domestic animals.

Classification of epigenetic modifications

Epigenetic modifications such as DNA methylation, histone modifications, and non-coding RNAs can affect the expression patterns of genes related to economic traits in livestock, thereby exerting significant effects on livestock performance. For

example, Liu et al. found that the expression level of miR-499-5p was significantly positively correlated with the content of skeletal muscle type I fibers and significantly negatively correlated with the content of type IIB fibers by comparing the gene expression patterns of different types of chicken muscle fibers. This study suggests that non-coding RNA is a type of epigenetic modification that exists in different quality chicken breeds with differential expression levels. In addition, Caiye et al. conducted a genome-wide DNA methylation association analysis on three breeds of sheep, Lanzhou Big-Tail Sheep, Altay Sheep, and Tibetan Sheep, and identified differential methylated regions or sites, revealing the differential DNA methylation modifications during tail formation in different sheep breeds. Meanwhile, Erven et al. used interpolation with low coverage ancient *Sus scrofa* data to detect genetic mutations and polymorphisms resulting from the evolution of ancient DNA (aDNA), and inferred errors in identifying or over-representing ancestral components and selective features in the genotype. These findings shed light on the mechanisms behind the emergence and spread of early animal husbandry in Europe and the Near East during the entire Neolithic era, based on genetic variation.

Genetic markers associated with economic traits in domestic animals

Focusing on genetic markers associated with novel economic traits in livestock can help to screen for individuals carrying superior genetic markers, thereby contributing to the productivity and economic efficiency of livestock. Wen et al. compared the spectra of extrachromosomal circular DNA (eccDNA) and extrachromosomal DNA (ecDNA) between Wei pigs (WP) and Large White pigs (LP), and found that the abundance of eccDNA was lower, while the abundance of ecDNA was higher in WP compared to LP. Although eccDNA and ecDNA are common in both WP and LP, their sizes are large enough to carry one or more partial or complete genes, making them a potential new genetic marker for selecting high-quality breeds. A study by Cao et al. found that acyl-CoA synthetase long-chain family member 1 (ACSL1) has two different transcripts, namely, ACSL1-a and ACSL1-b. Their research comprehensively elucidated the crucial role of the ACSL1 gene in triglyceride synthesis, which may provide molecular markers for screening high-quality lamb meat. Interestingly, because Alvarenga et al. found that temperament is highly heritable and polygenic, with favorable genetic correlations with other relevant traits, so the temperament of cattle has been regarded as a key breeding goal by farmers.

The relationship between epigenomics and economic traits

Epigenetic modifications play an important role in the economic traits of domestic animals, and in-depth research on the relationship between epigenetic modifications and economic traits can provide theoretical and practical guidance for domestic animals breeding and improvement of economic traits. Caiye et al. identified 68,603 differentially methylated regions (DMCs) and 75 differentially methylated genes (DMGs) associated with these DMCs, of which the DMGs (NFATC4, LPIN2, MGAT2, and MAT2B) are mainly involved in the process of fat metabolism, which providing the first evidence for the relationship between tail fat deposition in local sheep and epigenetics. Two studies by Zheng et al. revealed the dynamic changes in protein expression during the development of the rumen epithelial cell, identifying the major proteins involved in rumen epithelial cell development. These studies also provided preliminary evidence for the importance of post-translational modifications of histones in the development of the rumen epithelial cell. It is necessary to screen epigenetic modifications that regulated DMGs associated with economic traits for further application in breeding.

Prospects for improving economic traits based on epigenetic modifications

Using epigenetic modification techniques, the genome can be directly edited and regulated, allowing for rapid changes in the economic traits in domestic animals. Wang et al. reported that transgenic mice were generated, which exhibit skin-specific overexpression of ovarian β-catenin. The excess expression of β-catenin promotes the transition of hair follicles from the growth phase to the maturation phase, thereby increasing hair follicle density. This provides a new approach for using transgenic technology to improve economic traits in domestic animals. Different economic traits may be associated with different epigenetic modifications. By regulating these epigenetic modifications, simultaneous improvement of multiple traits can be achieved. In fact, Liu et al. found that non-coding RNAs are associated with chicken meat quality, while Caiye et al. suggested that DNA methylation is related to the formation of sheep tail fat, which confirms the association between different traits and different genetic markers. The prospects for improving economic traits based on epigenetic modifications are broad, but breeding methods based on epigenetic modifications are still in the developmental stage and require further research and validation.

In summary, this Research Topic collected a large amount of research evidence confirming the close relationship between epigenetic modifications and economic traits in domestic animals, while proposing new breeding ideas for improving animal economic benefits.

Author contributions

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

Funding

The work was supported in part by the National Natural Science Foundation of China (32002160), the Anhui Province Key Research and Development Program Project (202204c06020074), Anhui Provincial Natural Science Foundation (2008085QC140), the University Research Project of Anhui Province (2022AH040232 and KJ2020A008), Tianchang City Intelligent Equipment and Instrumentation Research Institute Special Funding Project (tzy202213), the Foundation of Anhui Science and Technology University (DKYJ201901), the Graduate Program of the Anhui Provincial Department of Education (YJS20210557), and the Innovation Funds for Undergraduate Students of Anhui Province (202110879059, S202110879175, S202210879200, and S202210879194).

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

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 04 April 2022
doi: 10.3389/fgene.2022.794625

Genetic Modeling and Genomic Analyses of Yearling Temperament in American Angus Cattle and Its Relationship With Productive Efficiency and Resilience Traits

*Amanda B. Alvarenga[1], Hinayah R. Oliveira[1,2], Stephen P. Miller[3], Fabyano F. Silva[4] and Luiz F. Brito[1]**

[1]Department of Animal Sciences, Purdue University, West Lafayette, IN, United States, [2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, [3]American Angus Association, Angus Genetics Inc., St Joseph, MO, United States, [4]Department of Animal Sciences, Federal University of Vicosa, Viçosa, Brazil

OPEN ACCESS

Edited by:
Fernando Baldi,
São Paulo State University, Brazil

Reviewed by:
Just Jensen,
Aarhus University, Denmark
Ludmilla Costa Brunes,
Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Angelica Pereira,
University of São Paulo, Brazil

***Correspondence:**
Luiz F. Brito
britol@purdue.edu

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

Received: *13 October 2021*
Accepted: *25 February 2022*
Published: *04 April 2022*

Citation:
Alvarenga AB, Oliveira HR, Miller SP, Silva FF and Brito LF (2022) Genetic Modeling and Genomic Analyses of Yearling Temperament in American Angus Cattle and Its Relationship With Productive Efficiency and Resilience Traits. Front. Genet. 13:794625. doi: 10.3389/fgene.2022.794625

Cattle temperament has been considered by farmers as a key breeding goal due to its relevance for cattlemen's safety, animal welfare, resilience, and longevity and its association with many economically important traits (e.g., production and meat quality). The definition of proper statistical models, accurate variance component estimates, and knowledge on the genetic background of the indicator trait evaluated are of great importance for accurately predicting the genetic merit of breeding animals. Therefore, 266,029 American Angus cattle with yearling temperament records (1–6 score) were used to evaluate statistical models and estimate variance components; investigate the association of sex and farm management with temperament; assess the weighted correlation of estimated breeding values for temperament and productive, reproductive efficiency and resilience traits; and perform a weighted single-step genome-wide association analysis using 69,559 animals genotyped for 54,609 single-nucleotide polymorphisms. Sex and extrinsic factors were significantly associated with temperament, including conception type, age of dam, birth season, and additional animal–human interactions. Similar results were observed among models including only the direct additive genetic effect and when adding other maternal effects. Estimated heritability of temperament was equal to 0.39 on the liability scale. Favorable genetic correlations were observed between temperament and other relevant traits, including growth, feed efficiency, meat quality, and reproductive traits. The highest approximated genetic correlations were observed between temperament and growth traits (weaning weight, 0.28; yearling weight, 0.28). Altogether, we identified 11 genomic regions, located across nine chromosomes including BTAX, explaining 3.33% of the total additive genetic variance. The candidate genes identified were enriched in pathways related to vision, which could be associated with reception of stimulus and/or cognitive abilities. This study encompasses large and diverse phenotypic, genomic, and pedigree datasets of US Angus cattle. Yearling temperament is a highly heritable and polygenic trait that can be improved through genetic selection. Direct selection for temperament is not expected to result in

unfavorable responses on other relevant traits due to the favorable or low genetic correlations observed. In summary, this study contributes to a better understanding of the impact of maternal effects, extrinsic factors, and various genomic regions associated with yearling temperament in North American Angus cattle.

Keywords: behavior, genome-wide association study, livestock, long noncoding RNA genes, xlinked trait, weighted single-step GBLUP

INTRODUCTION

Genetic evaluation schemes have been refined over time concomitantly with consumers, producers, and industry requirements. A recent American Angus Association (AAA) producers' survey revealed docility (or temperament) as one of the top three traits to be prioritized in US Angus cattle breeding programs (Santos et al., 2019). Behavioral traits are important due to their impact on animal performance [e.g., growth (Hoppe et al., 2010)], meat quality (King et al., 2006; Coutinho et al., 2017; Sant'Anna et al., 2019), reproduction (Cooke et al., 2011; Valente et al., 2015; Cziszter et al., 2016), and immunity (Burdick et al., 2011). Other substantial influences include handler safety, animal welfare, longevity (Neja et al., 2015), efficiency of management systems, and economic profit (Carlstrom et al., 2014). For instance, Busby (2015) has reported an average profit of $46.63 per docile animal compared to $7.62 per animal with aggressive behavior. Temperament is one of a plethora of behavioral measurements, and it is defined as "the animal response to handling or forced movement by humans" (Burrow et al., 1988). Temperament can be assessed in multiple ways, including objective measurements [e.g., flight time (Burrow et al., 1988)] and subjective evaluations (Tulloh, 1961; Dickson et al., 1970; Kilgour, 1975).

A wide range of heritability estimates [average equal to 0.27, ranging from 0.03 to 0.67 (Sato, 1981; Hearnshaw and Morris, 1984; Hoppe et al., 2010; Rolfe et al., 2011)] has been reported for beef cattle temperament traits, supporting the idea that cattle temperament can be improved through genetic or genomic selection. Studies underlying the genetic background of temperament have also been performed (Esmailizadeh et al., 2008; Chan, 2012; Hanna et al., 2014; Riley et al., 2016; Garza-Brenner et al., 2017; Macleod et al., 2019; Costilla et al., 2020). In total, 29 quantitative trait loci (QTL) have been reported, in which 19 were identified for temperament in beef or dual-purpose cattle (Esmailizadeh et al., 2008; Glenske et al., 2011; Riley et al., 2016; Dos Santos et al., 2017; Garza-Brenner et al., 2017; Hu et al., 2019). Additionally, orthologous genes have been reported to be controlling behavioral indicator traits across various species, including humans (Costilla et al., 2020; Alvarenga et al., 2021). However, to the best of our knowledge, none of the mentioned studies have used large datasets [the maximum number of records in previous studies was 9,223 individuals (Costilla et al., 2020)] recorded in a broad array of geographic, environmental, and climatic zones, nor genetically evaluated temperament in the US Angus population, which is the most predominant breed in the largest producer of beef cattle in the world (Greenwood, 2021). Along with the breeding goals, environmental conditions, management practices, genetic population parameters (Hidalgo et al., 2020), and trait definitions change over time. Herein, routine updates of the statistical models and re-estimation of variance components are required, as the accuracy of estimated breeding values (EBVs) relies on properly defined genetic models and accurate genetic parameter estimates.

Therefore, our overarching goal was to better understand the variation of temperament at the yearling age across the US using a large dataset provided by the AAA through the Angus Genetics Inc. (AGI—American Angus Association; Saint Joseph, MO, US), from the non-genetic, genetic, and genomic aspects. The understanding of the non-genetic factors impacting temperament can guide the statistical modeling and provide a better understanding of nurture impacting temperament. Thereafter, up-to-date statistical models and variance components will allow more accurate EBV estimates and a more powerful genome-wide association to understand the genetic background and the biological mechanisms associated with cattle temperament. With that, the specific objectives of this study were to 1) identify the non-genetic effects (e.g., management events) influencing yearling temperament; 2) define an optimal statistical model (in terms of systematic and random effects) to be used in routine genetic and genomic evaluations of yearling temperament in North American Angus cattle; 3) estimate variance components for yearling temperament; 4) estimate approximated genetic correlations between yearling temperament and 20 productive efficiency and resilience traits, including calving ease, growth traits, carcass measurements, body conformation, susceptibility to high altitude disease, hair shedding score, and other mature cow measurements; and 5) identify genomic regions and candidate genes controlling yearling temperament in Angus cattle based on weighted single-step genome-wide association studies, followed by metabolic pathway (functional enrichment) analyses and a comparison with previously reported QTLs.

MATERIALS AND METHODS

Data

All the data used in this study are from purebred and commercial Angus cattle registered in the AAA, born between 1990 and 2018, and recorded for temperament score at yearling age. Yearling temperament score is a subjective measurement (i.e., visually evaluated by the farmers/handlers), and it is recorded from 320 to 440 days of age (yearling) when the animal exits a chute, which is consistent within contemporary groups (Northcutt and Bowman, 2007). Temperament is scored on a one-to-six scale, in which a

TABLE 1 | Description[a] of yearling temperament scores and number of animals per level in American Angus cattle.

Score	Description[a]	N
1	Docile—mild disposition. Gentle and easily handled. Stands and moves slowly during processing. Exits chute calmly	191,402
2	Restless—quieter than average but may be stubborn during processing. May try to back out of chute. Some flicking of tail. Exits chute promptly	58,927
3	Nervous—typical temperament is manageable, but nervous and impatient. Displays a moderate amount of struggling, movement, and tail flicking. Exits chute briskly	13,615
4	Flighty (wild)—jumpy and out of control, quivers, and struggles violently. May bellow and froth at the mouth. Displays continuous tail flicking. Defecates and urinates during processing. May jump when penned individually. Exhibits long flight distance and exits chute wildly	1,778
5	Aggressive—may be similar to score 4, but added aggressive behavior, fearfulness, extreme agitation, and continuous movement, which may include jumping and bellowing while in the chute. Exits chute frantically and may exhibit attack behavior when handled alone	235
6	Very aggressive—thrashes about or attacks wildly when confined in small, tight places. Pronounced attack behavior	72

[a]*Description and scoring guidelines from the ANGUS Journal Report, October 2007 (Northcutt and Bowman, 2007); N: Number of animals per score of yearling temperament after the data quality control.*

score of 1 represents a docile animal (i.e., desirable behavior) and a score of 6 represents a very aggressive (more temperamental) animal. A complete description of the scores and measurement guidelines used by the AAA are shown in **Table 1**.

Records from 675,678 Angus cattle with temperament score at yearling age were available. Few animals had scores of 4 (1,778 animals), 5 (425 animals), or 6 (151 animals). Therefore, these three categories were combined, which resulted in 83.8% of animals with score 1, 12.9% animals with score 2, 2.8% animals with score 3, and 0.5% animals with score 4+ (representing the scores 4, 5, and 6). A quality control procedure was performed to remove animals with missing information for the systematic effects evaluated and to warranty enough variability within their levels (i.e., at least three animals per level). Thereafter, the herds (where animals were located at birth, weaning, and yearling) exclusively represented by one score (e.g., all animals with score 1) were removed from the dataset. This criterion was used to avoid bias generated by the farmer/evaluator and allow variability within the levels. Additionally, data from animals younger than 320 days and older than 440 days at the time of temperament assessment were also removed. After quality control, there were 266,029 animals comprising 147,671 bulls, 3,332 steers, and 115,026 females with yearling temperament scores 1 (71.9%), 2 (22.2%), 3 (5.1%), or 4+ (0.8%). These remaining animals were born between 2001 and 2018. The number of animals per yearling temperament score after the quality control is shown in **Table 1**.

Statistical Model Definition and Genetic Parameter Estimation

Before consideration of the statistical model to be used, several options were tested for environmental (systematic) and genetic (direct and maternal genetic) effects. The statistical models used in this study were defined based on two steps: 1) definition of the systematic effects and contemporary group (*CG*) and 2) definition of direct genetic, maternal genetic, and/or maternal environmental effects. Concomitantly to the second step, the genetic parameters were estimated.

TABLE 2 | Random effects included in the four animal models for yearling temperament in American Angus cattle.

Models	Components				
	σ^2_{CG}	σ^2_u	σ^2_m	$\sigma_{u,m}$	σ^2_p
D	✓	✓	—	—	—
DMG	✓	✓	✓	✓	—
DMP	✓	✓	—	—	✓
DMGP	✓	✓	✓	✓	✓

D: reduced model fitting the additive direct genetic effect; DMG: model fitting the additive direct and maternal genetic effects; DMP: model fitting the additive direct genetic and maternal environmental effects; DMGP: complete model including additive direct genetic, maternal genetic, and maternal environmental effects; σ^2_{CG}: contemporary group variance; σ^2_u: additive direct genetic variance; σ^2_m: maternal genetic variance; $\sigma_{u,m}$: covariance between additive direct and maternal genetic effects; σ^2_p: maternal environmental variance; ✓: component was included in the model.

The first step was performed using the *glm* function available in R software (R Core Team, 2019), considering a quasi-binomial family distribution. Stepwise (i.e., backward elimination) subroutines for selecting the best subset of covariates were used for the systematic model definition and to define the *CG*. The best fitted non-genetic model for yearling temperament rendered a coefficient of determination equal to 0.35. The model included age of dam (*AOD*; 10 levels), *CG* (22,322 levels), and conception type (*ET*; two levels: embryo transference or natural conception) as categorical systematic effects and calf age deviation from 365 days (*CALFDEV*) as a linear covariate. *AOD* was classified following the Beef Improvement Federation Guidelines [BIF (Cundiff et al., 2018)]. *CG* was defined as the concatenation of birth date (month-year), birth herd, weaning date (month-year), weaning herd, if the animal was submitted to a creep feeding system, if the animal had ultrasound information, and date (month-year), herd, sex, and group age deviation when yearling temperament was recorded.

After defining the systematic model and the optimal *CG*, four mixed animal models were tested by incorporating four random effects: *CG*, additive direct genetic, maternal genetic, and/or maternal environmental effects. The four models and their corresponding components are listed in **Table 2**.

Bayesian threshold models based on the Gibbs sampler and Markov Chain Monte Carlo (MCMC) algorithm were used to estimate the genetic parameters and, thereafter, compare the models. thrgibbs1f90 software (Tsuruta and Misztal, 2006) was used with a chain length varying from 500,000 to 1,000,000 iterations, assuming a burn-in from 50% to 75%, and a thin of 50 or 100. Non-informative prior information based on the inverse-chi-squared distribution was assumed for variance components. The models are presented below. The reduced model is defined as follows:

$$\text{D model: } \mathbf{l} = \mathbf{Xb} + \mathbf{Ww} + \mathbf{Zu} + \mathbf{e} \quad [1]$$

where $\mathbf{l}$ is a vector of observations for the underlying threshold trait measured on the liability scale for yearling temperament score; $\mathbf{b}$ is a vector of systematic effects, including *AOD* and *ET* and *CALFDEV* as a linear covariate, $\mathbf{b}$ assumes a uniform distribution [non-informative prior, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{\Sigma_b} \otimes \mathbf{I})$]; $\mathbf{w}$ is a vector of CG random effects, *CG*, $\mathbf{w} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{\mathrm{w}})$; $\mathbf{u}$ is a vector of direct genetic effect, $u \sim N(\mathbf{0}, \mathbf{A}\sigma^2_{\mathrm{u}})$; e is a residual vector, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{\mathrm{e}})$; $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{Z}$ are incidence matrices of $\mathbf{b}$, $\boldsymbol{w}$, and $\mathbf{u}$, respectively. The $\mathbf{\Sigma}_b$ is a diagonal matrix with large values to represent vague prior knowledge, $\mathbf{A}$ is the pedigree-based relationship matrix assuming 10 generations back to the animals with phenotypes, $\mathbf{I}$ is an identity matrix, and $\otimes$ denotes the Kronecker product. σ^2_w, σ^2_u, and σ^2_e are the *CG*, additive genetic, and residual variances, respectively. The three models incorporating maternal effects are as follows:

$$\text{DMG model: } \mathrm{l} = \mathbf{Xb} + \mathbf{Ww} + \mathbf{Zu} + \mathbf{Z_2m} + \mathbf{e} \quad [2]$$

where $\mathbf{u}$ is a vector of direct genetic effect on the liability scale; $\mathbf{m}$ is the maternal genetic effect; $\mathbf{Z_2}$ is an incidence matrix of $\mathbf{m}$, $\mathbf{m}|\sigma^2_m \sim N(\mathbf{0}, \mathbf{A}\sigma^2_{\mathrm{m}})$; the covariance structure of genetic random effects ($\mathbf{u}$ and $\mathbf{m}$) was defined as $\mathbf{A} \otimes \sum$, in which $\mathbf{\Sigma} = \begin{bmatrix} \sigma^2_u & \sigma_{u,m} \\ \sigma_{u,m} & \sigma^2_m \end{bmatrix}$; σ^2_m and $\sigma_{u,m}$ are the maternal genetic variance and direct and maternal genetic covariance, respectively. All other terms were previously defined in **Eq. 1**.

$$\text{DMP model: } \mathbf{l} = \mathbf{Xb} + \mathbf{Ww} + \mathbf{Zu} + \mathbf{Sp} + \mathbf{e} \quad [3]$$

where $\mathbf{p}$ is a vector of maternal environmental effect, $\mathbf{p}|\sigma^2_p \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{\mathrm{p}})$; S is an incidence matrix of $\mathbf{p}$, and σ^2_p is the maternal environmental variance. All other terms were previously defined in **Eq. 1**.

$$\text{DMPG model: } \mathbf{l} = \mathbf{Xb} + \mathbf{Ww} + \mathbf{Zu} + \mathbf{Z_2m} + \mathbf{Sp} + \mathbf{e} \quad [4]$$

where all the components of the DMGP model were previously defined in **Eqs 1–3**. The systematic effects were the same for all four models.

A total of 819,303 animals (242,570 bulls, 298 steers, and 576,435 cows) born between 1928 and 2018 were included in the *A* matrix. Of those, 775,176 animals had both known parents, 35 animals had one known parent, and 44,092 animals had unknown parents. The MCMC convergence was verified using the Heidelberger and Welch (Heidelberger and Welch, 1983) and Geweke (Geweke, 1991) criteria, both implemented in the *boa* package (Smith, 2007) of R software (R Core Team, 2019). The variance components and heritability estimates were obtained using postgibbsf90 software (Tsuruta and Misztal, 2006). The direct (h^2_d), maternal (h^2_m), and total (h^2_t) heritability estimates were computed, respectively, as follows:

$$h^2_d = \frac{\sigma^2_u}{\sigma^2_t} \quad [5]$$

$$h^2_m = \frac{\sigma^2_m}{\sigma^2_t} \quad [6]$$

$$h^2_t = \frac{\sigma^2_u + 1.5\sigma_{u,m} + 0.5\sigma^2_m}{\sigma^2_t} \quad [7]$$

where σ^2_t is the total variance (which was differently calculated for each model; i.e., σ^2_t is equal to the sum of the σ^2_u, σ^2_w, and σ^2_e elements for the D model; the sum of σ^2_u, σ^2_w, $\sigma_{u,m}$, σ^2_m, and σ^2_e for the DMG model; the sum of σ^2_u, σ^2_w, σ^2_p, and σ^2_e for the DMP model; and the sum of σ^2_u, σ^2_w, $\sigma_{u,m}$, σ^2_m, σ^2_p, and σ^2_e for the DMGP model).

The models were compared based on the linear regression method (Legarra and Reverter, 2018), which estimates bias, dispersion, and accuracies comparing two subsets of breeding values (EBVs) estimated using a partial and whole dataset. The whole dataset comprises all animals with temperament records (i.e., 266,029). The partial dataset was defined as animals in the whole dataset but masking the phenotype of the animals born in 2018 (13,202 animals), which was defined as the validation group. The EBVs from the partial and whole datasets were obtained fixing the variance components in thrgibbs1f90 with 10,000 iterations, 1,000 burn-in, and 5 thin. Bias, dispersion, and predictive accuracies were calculated for the validation animals.

Effect of Sex and Extrinsic Variables on Yearling Temperament

Sex and extrinsic factors available in the AAA dataset were submitted to a pairwise comparison using the *lsmeans* package (Lenth, 2016) implemented in R software (R Core Team, 2019). A linear model including calf age deviation from 365 days as a linear covariate and the following target independent variables as a categorical fixed effect, such as age of dam, conception type (embryo transfer or natural conception), parity type (single or twin), sex, birth season, if the animal was under a creep-feeding system, and if the animal had ultrasound and feed intake information, was used. In other words, the mean presented for each level of the fixed effect was adjusted for the remaining factors.

Approximated Genetic Correlations

The genetic relationship between yearling temperament and other productive efficiency and resilience traits is of great importance for the design of breeding programs and development of selection indexes. Due to limited access to the raw datasets and computational limitations, the genetic correlation between yearling temperament and 20 other relevant traits (**Table 3**) was assessed based on the correlation between the EBVs. The EBVs for yearling temperament were

TABLE 3 | Description of production traits evaluated to be genetically correlated with yearling temperament score, as defined by the American Angus Association.

Symbol	Trait EBV	Unit and description[a]
CED	Calving ease-direct	Percentage of unassisted births (calf measurement/direct genetic effect)
BW	Birth weight	Pounds
WW	Weaning weight	Pounds, direct genetic effect
YW	Yearling weight	Pounds
RADG	Residual average daily gain	Pounds per day, the sire's ability for post-weaning gain in his progeny given a constant amount of feed consumed
DMI	Dry-matter intake	Pounds per day
YH	Yearling height	Inches
SC	Scrotal circumference	Centimeters
PAP	Pulmonary artery pressure	Probability, the sire's ability to produce a progeny with lower (or greater) pulmonary arterial pressures probability, decreasing (or increasing) the risk of contracting high altitude diseases
HP	Heifer pregnancy	Percentage, it measures the ability of the sire's daughters to become pregnant as first-calf heifers during a normal breeding season
CEM	Calving ease maternal	Percentage of unassisted births (cow measurement/maternal genetic effect)
MILK	Maternal milk	Pounds of calf weaned, the sire's genetic merit for milk and mothering ability as expressed in his daughters (maternal genetic effect of WW)
MW	Mature weight	Pounds
MH	Mature height	Inches
CW	Carcass weight	Pounds
MARB	Marbling score	Marbling score
RE	Ribeye area	Square inches
FAT	Fat thickness	Inches, the sire's ability to transmit fat thickness at the 12th rib (as measured between the 12th and 13th ribs) to his progeny
FOOT	Foot angle score	Foot-angle score, the sire's ability to transmit ideal foot angle to his progeny, of which a lower value is desirable
HS	Hair shed score	Hair shed score, the sire's ability to transmit early (or late) summer hair shedding, of which a lower value is desirable

EBV: Estimated breeding value.

[a]*Source: American Angus Association website (www.angus.org/mobile/nce/definitions.aspx).*

from the *Statistical Model Definition and Genetic Parameter Estimation* while fixing the variance components in thrgibbs1f90. The EBVs for the other traits were from the official genetic evaluation performed by the AAA, which includes the following: calving ease [calving ease direct (CED) and maternal calving ease (CEM)], body weight [birth weight (BW), weaning weight (WW), and yearling weight (YW)], residual average daily gain (RADG), dry-matter intake (DMI), yearling height (YH), scrotal circumference (SC), carcass and meat quality measurements [carcass weight (CW), marbling score (MARB), ribeye area (REA), and fat thickness (FAT)], conformation [foot angle (FOOT)], adaptation [high altitude disease susceptibility—pulmonary artery pressure (PAP) and hair shedding score (HS)], and other mature cow measurements [fertility—heifer pregnancy (HP), mature weight (MW), mature height (MH), and maternal milk (MILK)]. A complete description of the traits is shown in **Table 3**. Data from animals with EBV accuracies lower than 0.25 for both evaluated traits were disregarded for the correlation. A weighted Pearson correlation ($r_{w1,2}$) between EBV pairs was calculated. The calculation of $r_{w1,2}$ comprised five steps:

1st. Calculation of the weights (w_i), which was based on the accuracies of the respective traits to be correlated for the animal *i*:

$$w_i = \frac{\mathrm{REL_{1i}}*\mathrm{REL_{2i}}}{\sqrt{\mathrm{REL_{1i}}*\mathrm{REL_{2i}}}} \quad [8]$$

where w_i is the weighting for the animal *i*; REL_{1i} and REL_{2i} are the squared accuracies of the EBV for traits 1 and 2 for animal *i*, respectively. The approximate squared accuracies (REL_1 and REL_2) of the EBV was calculated using accf90GS software (Misztal et al., 2013).

2nd. Calculation of the weighted mean for traits 1 and 2:

$$m_1 = \frac{\sum_{i=1}^{n}(\mathrm{EBV_{1i}}*\mathrm{w_i})}{\sum_{i=1}^{n}\mathrm{w_i}} \quad and \quad m_2 = \frac{\sum_{i=1}^{n}(\mathrm{EBV_{2i}}*\mathrm{w_i})}{\sum_{i=1}^{n}\mathrm{w_i}} \quad [9]$$

where m_1 and m_2 are the weighted mean for traits 1 and 2; EBV_{1i} and EBV_{2i} are the EBVs for traits 1 and 2, respectively, both measured in animal *i*, and *n* is the total number of animals. The EBV for yearling temperament was converted to the probability of the animal being docile (score 1), in which greater EBVs represent higher probability of the animal receiving a score of 1 (docile).

3rd. Calculation of the weighted variances:

$$s_1 = \frac{\sum_{i=1}^{n}\mathrm{w_i}(\mathrm{EBV_{1i}} - m_1)^2}{\sum_{i=1}^{n}\mathrm{w_i}} \quad and \quad s_2 = \frac{\sum_{i=1}^{n}\mathrm{w_i}(\mathrm{EBV_{2i}} - m_2)^2}{\sum_{i=1}^{n}\mathrm{w_i}} \quad [10]$$

where s_1 and s_2 are the weighted variance for traits 1 and 2 and all the remaining parameters have been previously described.

4th. Calculation of the weighted covariance:

$$s_{1,2} = \frac{\sum_{i=1}^{n}\mathrm{w_i}(\mathrm{EBV_{1i}} - m_1)(\mathrm{EBV_{2i}} - m_2)}{\sum_{i=1}^{n}\mathrm{w_i}} \quad [11]$$

where $s_{1,2}$ is the weighted covariance between EBVs for traits 1 and 2 and all the remaining parameters have been previously described.

5th. Calculation of weighted correlation:

$$r_{w1,2} = \frac{s_{1,2}}{\sqrt{s_1 * s_2}} \quad [12]$$

where $r_{w1,2}$ is the weighted correlation between EBVs for traits 1 and 2 and all the remaining parameters have been previously described.

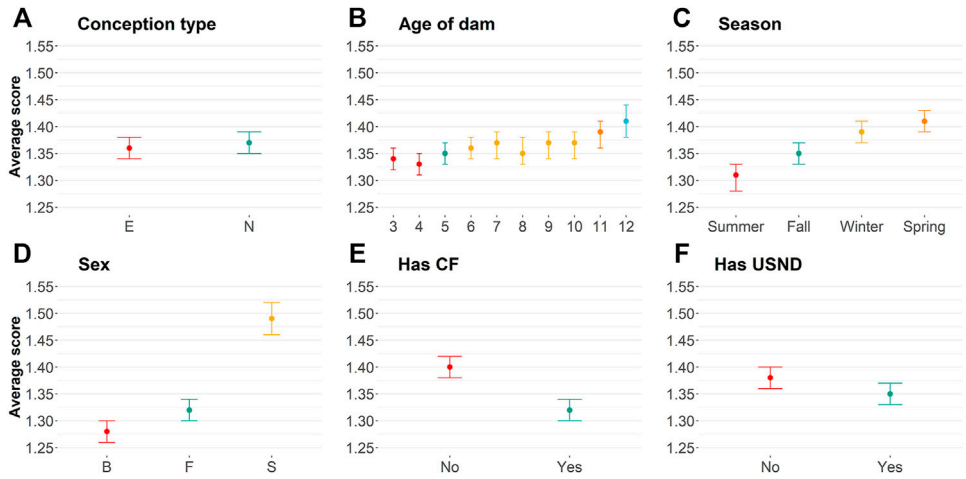


FIGURE 1 | Average of temperament score using least-square means for **(A)** conception type [naturally conceived (N) or embryo transference (E)]; **(B)** Age of the dam in years; **(C)**: birth season; **(D)**: sex [bull (B), steers (S), and female (F)]; **(E)** if the animal was creep-fed; and **(F)**: if the animal had ultrasound information.

Finally, the standard errors (± SE) of the genetic correlation metric (i.e., $r_{w1,2}$) were calculated as follows:

$$SE = \sqrt{\frac{1 - r_{w1,2}}{n - 2}} \quad [13]$$

where all the parameters have been described. The 95% confidence intervals were obtained as mean ± 1.96 × SE (Altman and Bland, 2005).

Weighted Single-Step Genome-Wide Association Analyses

A weighted single-step genome-wide association study [WssGWAS (Wang et al., 2012)] was performed to identify candidate regions controlling yearling temperament. From the 266,029 animals with phenotypic information for yearling temperament, 69,559 were genotyped. Animals were genotyped with various SNP arrays as part of ongoing commercial genotyping activities by breeders for genetic evaluation purposes, resulting in a market set of 54,609 SNPs. Commercial genotyping products were from Zoetis, including i50K (www.zoetisus.com/animal-genetics/media/documents/i50k-00001_50k-sellsheet.pdf) and HD50K (www.zoetisus.com/animal-genetics/beef/hd-50k/hd-50k-for-black-angus.aspx), and Neogen GeneSeek, including GeneSeek Genomic Profile Low-Density (GGPLD; 40 K SNPs), High-Density (GGPHD; 80 K SNPs), GGPUHD (150 K SNPs), and AngusGS. Both Zoetis and Neogen provided, for genomic evaluation purposes, an imputed (average and standard deviation imputation accuracy equal to 99.72% and 0.87) SNP set similar to the Illumina BovineSNP50V2 and Illumina BovineSNP50V3 (Illumina, Inc., San Diego, CA), respectively, and mapped to the bovine genome assembly UMD3.1. However, the genomic coordinates were converted to ARS-UCD1.2 bovine genome assembly (Medrano, 2017; Rosen et al., 2018) using the *biomaRt* R package (Durinck et al., 2009).

Quality control (QC) procedures were applied to remove genotyped individuals with a call rate lower than 90% and pedigree errors. SNP genotypes with a call rate lower than 90%, an MAF lower than 0.01, and a deviance of heterozygous genotype from the Hardy Weinberg Equilibrium higher than 0.15 were also removed. Both autosomal chromosomes and pseudo-autosomal regions of the X chromosome were used in the WssGWAS. SNPs located on approximated pseudo-autosomal regions [PAR; SNP with a genomic coordinate above X:133,300,518 bp; 109 SNPs; (Johnson et al., 2019)] represented the X chromosome because it follows an autosomal-chromosome inherence and, therefore, were analyzed as autosomal SNPs (Johnson et al., 2019). preGSf90 software (Aguilar et al., 2014) was used to perform the QC, in which 69,441 genotyped animals and 42,662 SNPs remained for further analyses. As a WssGWAS was performed, the pedigree-based relationship matrix (**A**) and **G** were combined into the **H** matrix (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010; Wang et al., 2014). The **H** inverse ($\mathbf{H^{-1}}$) is defined as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (0.90\mathbf{G} + 0.10\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \quad [14]$$

where $\mathbf{A}_{22}$ is a pedigree-based relationship matrix for the genotyped animals. The pedigree dataset contained animals traced back up to four generations, resulting in 578,821 animals. The **G** matrix was constructed as in the study by VanRaden (2008):

$$\mathbf{G} = \mathbf{RDR}' \quad [15]$$

where **R** is a matrix of gene content adjusted for observed allele frequencies and **D** is a diagonal matrix of SNP weights. The weights were derived from the third iteration's SNP solutions using a nonlinear approach described by VanRaden (2008) and Cole et al. (2009).

TABLE 4 | Fitness of the model and genetic parameters for all models tested.

Model	σ_e^2	$\hat{\Delta}_v$	$\hat{b}_v$	$\widehat{acc}_v$	h_d^2	h_m^2	h_t^2	$r_{d,m}$
D	0.21 (0.02)	0.02	0.97	0.20	0.39 (0.01)			
DMG	0.19 (0.02)	0.00	0.95	0.18	0.44 (0.02)	0.04 (0.01)	0.38 (0.01)	−0.40 (0.04)
DMP	0.21 (0.02)	−0.02	0.96	0.21	0.38 (0.01)			

D: reduced model, including direct genetic effect; DMG: model including direct genetic and maternal genetic effects; DMP: model including direct genetic and maternal permanent environmental effects; σ_e^2: residual variance (standard error); $\hat{\Delta}_v$: bias; $\hat{b}_v$: dispersion; $\widehat{acc}_v$: predictive accuracy; h_d^2: direct genetic heritability (standard error); h_m^2: maternal genetic heritability (standard error); $r_{d,m}$: correlation between direct-genetic and maternal (standard error). The variance components are on the liability scale.

The optimal or selected statistical model was used to perform the WssGWAS. The vector of direct genetic effects was assumed to follow a normal distribution, $\mathbf{u}|\sigma_u^2 \sim N(\mathbf{0}, \mathbf{H}\sigma_u^2)$. The variance components were fixed based on the estimates given by the pedigree-based genetic analyses. thrgibbs1f90 software (Tsuruta and Misztal, 2006) with a 10,000 chain length, 1,000 burn-in, and 5 thin was used to calculate the genomic estimated breeding values (GEBVs). The SNP effects were obtained by back-solving the GEBVs using postGSf90 software (Stranden and Garrick, 2009; Wang et al., 2012; Aguilar et al., 2014). Sliding genomic windows of five SNPs were considered to calculate the effect of a certain genomic region on yearling temperament and the proportion of the total additive genetic variance explained. The window size was defined based on linkage disequilibrium (LD) decay in Angus cattle [LD higher than 0.19 (Lu et al., 2012)]. Genomic windows explaining more than 0.20% of the total additive genetic variance were considered as relevant and subjected to further investigations.

The genomic coordinates were based on the ARS-UCD1.2 bovine genome assembly (Medrano, 2017; Rosen et al., 2018). Quantitative trait loci (QTL) curated in the Cattle QTL DataBase [Cattle QTLdb (Hu et al., 2019); www.animalgenome.org; access date: September 27, 2021] and located within the selected genomic windows were identified. Gene annotation information was retrieved from Ensembl using the *biomaRt* R package (Durinck et al., 2009). Functional annotation was performed in terms of Gene Ontology (GO) biological processes [GO_BP; (Blake et al., 2013)] and metabolic pathways of the Kyoto Encyclopedia of Genes and Genomes [KEGG; (Kanehisa and Goto, 2000)] available in the DAVID database [david.ncifcrf.gov/tools.jsp; (Dennis et al., 2003); access date: September 27, 2021].

RESULTS

Sex and Extrinsic Variables Affecting Yearling Temperament

Yearling temperament was significantly associated with sex and extrinsic factors, which includes reproductive conception type, age of dam, birth season, creep-fed or non–creep-fed animals, and if there is ultrasound and/or feed intake recording (**Supplementary Table S1**). Parity type (i.e., single or twin) was the only factor not significantly associated with temperament.

Embryo-transferred animals were statistically more docile (average equal to 1.36) than naturally conceived animals (average equal to 1.37; **Supplementary Table S1**; **Figure 1A**). Small differences in the averages are probably due to a skewed distribution toward docile scores (**Table 1**). The age of dam categorized by years (from three to 12 years of age) also influenced yearling temperament (**Figure 1B**; **Table 2**; **Supplementary Table S1**). Dams up to 4 years old raised more docile progenies (average equal to 1.34) than older dams (average equal to 1.41).

Birth season was statistically associated with yearling temperament score ($p \leq 0.05$; **Supplementary Table S1**; **Figure 1C**). Animals that were born during the summer, fall, winter, and spring tend to range from docile to aggressive, respectively. The sex of the individual also influenced yearling temperament average. Bulls, females, and steers tended to be from more docile to more aggressive animals ($p < 0.05$; **Supplementary Table S1**; **Figure 1D**), respectively. Bulls had an average of yearling temperament score equal to 1.28, females had a temperament average equal to 1.32, and steers had a 1.49 average.

Our phenotypic analyses have also shown a positive relationship between implementation of the creep feeding system and animal temperament (**Supplementary Table S1**; **Figure 1E**). In short, creep-fed calves tend to be more docile (average of 1.32) than non–creep-fed calves (average of 1.40). Finally, animals that, in general, had additional recorded phenotypic collections were more docile ($p \leq 0.05$; **Supplementary Table S1**; **Figure 1F**). For instance, animals that had ultrasound information tended to be more docile (average of 1.35) than animals without this measurement (average of 1.38). The same pattern was observed for feed intake information, which was expected because 90% of animals that had feed intake also had ultrasound information (i.e., 1,362 had ultrasound and feed intake information out of 1,513 animals, while the rest had only feed intake information).

Model Choice and Parameter Estimation

We tested four animal models, which included direct genetic and maternal effects (**Table 2**). The results of the model incorporating both maternal effects (DMGP model including direct genetic, maternal genetic, and maternal environmental effects) are not presented because it did not converge even using up to 1 million iterations. The fitness measures of the models and genetic parameters are shown in **Table 4**. The DMG model, including direct and maternal genetic effect, provided a slight lower residual variance ($\sigma_e^2 = 0.19$), which is beneficial, compared to the other models ($\sigma_e^2 = 0.21$). The bias and dispersion from the forward validation were low for all models, varying from −0.02 to 0.02 and 0.95 and 0.97, respectively. Bias close to zero and dispersion close to one are desirable. Prediction accuracies were also similar across models, varying from 0.18 to 0.21. The computing time needed to

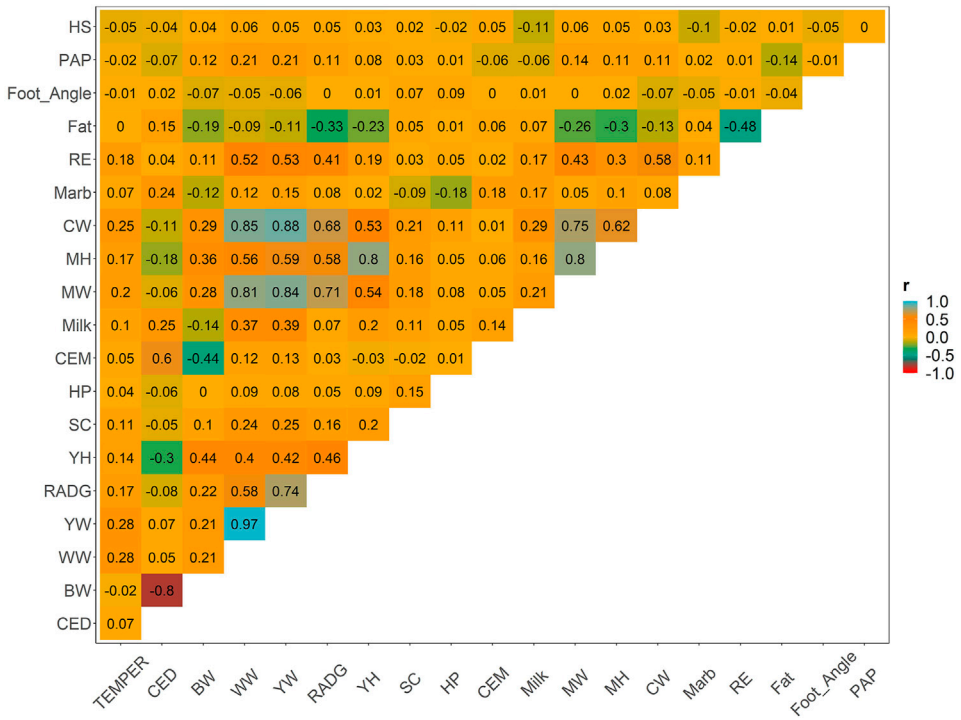


FIGURE 2 | Weighted Pearson correlation among all estimated breeding values for relevant traits in the beef cattle industry.

obtain EBVs using the whole dataset was 136.2, 239.5, and 240.4 minutes for the D, DMP, and DMG models, respectively.

Total yearling temperament heritability estimates ranged from 0.38 (DMP model) to 0.39 (D model) on the liability scale. Maternal genetic effects contributed to 4% of the total variation in yearling temperament on the liability scale (**Table 4**). Additionally, a strong negative genetic correlation was observed between direct and maternal genetic effects (−0.40; **Table 4**).

Approximated Genetic Correlation Between Yearling Temperament and Other Relevant Traits

Weighted correlations between the EBVs of yearling temperament and other key traits is shown in **Figure 2**. Additional correlations between all pairs of traits are provided for comparison. Other metrics, including the number of animals, and descriptive statistics (average, minimum, and maximum) of EBVs' theoretical accuracy are shown in **Supplementary Table S3**. Low-to-moderate genetic correlation was observed between temperament and all other relevant traits, which varied from −0.05 (yearling temperament vs. hair shedding) to 0.28 (yearling temperament vs. weaning or yearling weight). Other approximated genetic correlation between pairs of traits was similar to the genetic correlation estimated using multiple-trait models used by the AAA (please see www.angus.org/Nce/Heritabilities).

Positive and favorable association between yearling temperament and growth traits (i.e., direct weaning, yearling, and mature weight; correlation from 0.20 to 0.28), feed efficiency (i.e., RADG; 0.17), precocity (i.e., scrotal score; 0.11), and carcass traits (i.e., carcass weight, ribeye by area; varying from 0.18 to 0.25) were observed. The remaining traits had low genetic correlations, varying from −0.05 (i.e., hair shedding) to 0.10 (i.e., maternal weaning weight).

Weighted Single-Step Genome-Wide Association Analyses

The GEBVs from the D model were back-solved to generate the SNP window effects and the percentage of the total additive genetic variance explained by them. The WssGWAS was performed using 266,029 animals with phenotype, 69,441 animals with genotype for 42,662 SNPs located on autosomal and X chromosomes, and a pedigree containing 578,821 animals. The distribution of MAF is presented in **Supplementary Figure S1**, which was equally distributed across MAF intervals of 0.05.

Eleven independent genomic regions (i.e., non-overlapping regions of five SNPs) were identified, in which each genomic region explained more than 0.20% of the total additive genetic variance. All genomic regions together explained 3.33% of the total additive genetic variance for temperament. Relevant genomic regions were located on BTA2, BTA4, BTA8, BTA10, BTA11, BTA14, BTA26, BTA29, and X chromosomes (**Supplementary Table S4**). The Manhattan plot of the additive genetic variance explained by genomic windows for yearling temperament is shown in **Figure 3**.

Twenty-five positional candidate genes were identified surrounding the 11 genomic windows, in which 18 genes were

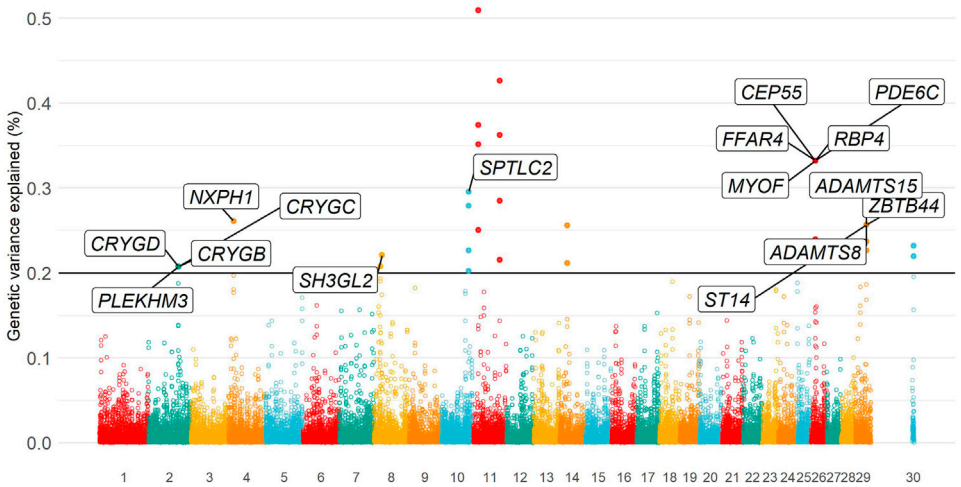


FIGURE 3 | Manhattan plot of additive genetic variance explained by genomic windows for yearling temperament using the D model.

TABLE 5 | Sample of the top genomic regions, the genes, and biological processes involved in these regions.

Gene name	Gene ensembl ID	CHR: start-end position	VE	Term
CRYGB	ENSBTAG00000048646	11:85006812-85223963	0.49	—
	ENSBTAG00000021770	2: 96181032-96426927	0.21	Eye lens protein
				Structural constituent of eye lens
CRYGC	ENSBTAG00000014783			Eye lens protein
				Methylation
				Visual perception
				Structural constituent of eye lens
CRYGD	ENSBTAG00000015054			Eye lens protein
				Visual perception
				Structural constituent of eye lens
PDE6C	ENSBTAG00000000445	26:14769909-14960555	0.33	Methylation
				Visual perception
U6	ENSBTAG00000042797	8: 26576536-26696264	0.22	—

CHR: Chromosome, VE: additive genetic variance explained by five-SNP window size (percentage).

annotated and their biotype was classified as protein coding (17 genes) and noncoding RNA genes (one gene; **Table 5**; **Supplementary Table S4**). The two regions with the largest variance explained (i.e., 0.50%) are located on BTA11, and no annotated gene has been observed in that region. However, one of the regions had one gene categorized as long noncoding RNA (lncRNA; **Supplementary Table S4**). The richest genomic regions for identified genes were BTA2, BTA26, and BTA29. The main genes annotated in these regions are *PLEKHM3*, *CRYGD*, *CRYGC*, *CRYGB*, *RBP4*, *MYOF*, *CEP55*, *FFAR4*, *PDE6C*, *ST14*, *ZBTB44*, *ADAMTS8*, and *ADAMTS15*.

There were 210 QTLs annotated in an overlap with the genomic windows associated with temperament. **Figure 4** shows the distribution of trait-type that these QTLs were previously associated with. In total, 72% of QTLs were previously associated with milk content, followed by 8% with structural problems, 6% with reproduction and production, and 4% with meat and carcass traits. Two QTLs for milking speed are located on a BTA14 genomic region explaining 0.26% of the total genetic variation. One gene overlapped in this region: *TOX* gene (thymocyte selection associated high mobility group box). In addition, two overlapped QTLs were annotated for longevity on BTA2, where four genes were identified, including three from the crystalline gamma family (i.e., *CRYGB, CRYGC*, and *CRYGD*) and *PLEKHM3* (pleckstrin homology domain containing M3).

To better understand the functions of the genes identified, we performed a functional analysis in DAVID, in which one GO biological process, one GO molecular function, and one term were significantly enriched based on the Benjamini test ($p < 0.10$). All terms were associated with vision processes. Visual perception (GO:0007601) was enriched by three genes located on BTA2, including *PDE6C* (phosphodiesterase 6C), *CRYGD*, and *CRYGC*. The last two genes from the *CRYG* family were also enriched in the structural constituent of eye lens (GO:0005212) and eye lens protein terms, in addition to a third CRYG family gene: *CRYGB* (**Supplementary Table S5**). Finally, another interesting process

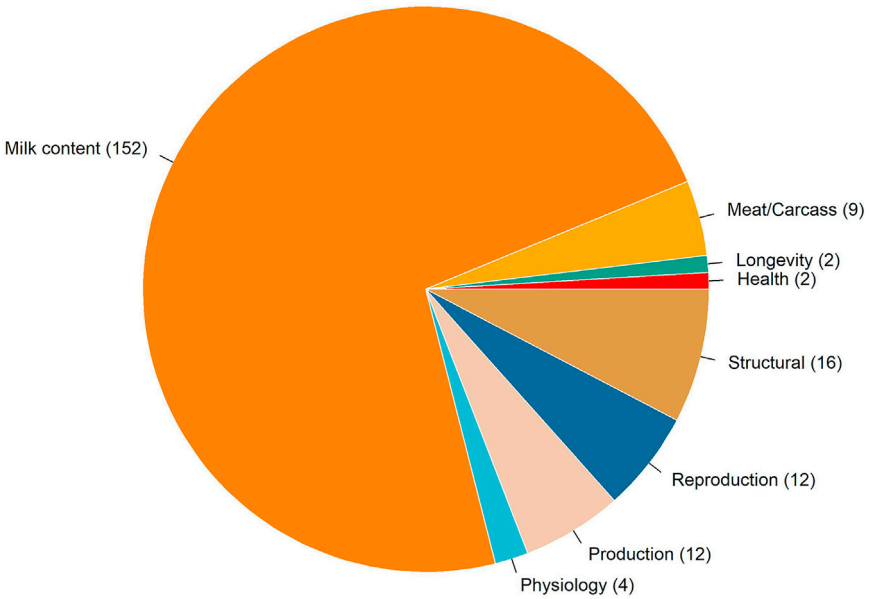


FIGURE 4 | Chart of the trait-types associated with the quantitative trait loci overlapping with the genomic regions associated with temperament.

that those genes also participate in is methylation (**Supplementary Table S5**). In **Table 5** is presented the genes enriched in the GO terms, as well as the genes located in the top regions.

DISCUSSION

Sex and Extrinsic Variables Affecting Yearling Temperament

Pointing out extrinsic factors influencing yearling temperament may inform future management studies and, hence, improve management strategies implemented in herds as well as handlers' safety. The elucidation of levels within the extrinsic factors intermeddling behavioral responses is also paramount to define and fit appropriate effects in the models, which can directly impact the prediction of EBVs and marker effects. Therefore, we have reported the connectivity between yearling temperament and farm experiences that the animal was subjected to, for which we used one of the largest beef cattle datasets available worldwide.

The statistical differences observed for embryo-transferred animals are potentially due to the genetic superiority of donors for many economically important traits, including temperament. An alternative hypothesis is associated with maternal ability and behavior during rearing. Particularly, foster/recipient cows could be (directly or indirectly) selected based on their maternal ability and easy-to-handle characteristics (Busby, 2015) in order to rear the biological progeny from high–genetic merit animals, and therefore, the behavior of these foster calves could be influenced by their foster mother characteristics.

Still, at the maternal level, younger dams raised more docile progenies than older dams. Similar trends were observed in Nellore cattle, where the average of temperament score tended to increase (i.e., being more aggressive) as the dam gets older (Barrozo et al., 2012). Studies have reported the high cognitive abilities of cattle, in which animals tend to avoid certain environments based on previous negative experiences (Ede et al., 2019). Thereupon, calves could reflect upon the behavior of their dams, in which an older cow, through memory acquisition, can become more aggressive or fearful across the years due to previous negative experiences (Kasimanickam et al., 2018).

In general, bulls have more dominant and aggressive behaviors than females (Reinhardt et al., 1987). Controversially, in recent studies, female beef cattle had been classified as more temperamental (Gauly et al., 2001). In this study, female Angus were on average more temperamental than male Angus ($p < 0.05$).

The birth-seasonal animal temperament pattern hinges on climate, grazing conditions, production system, and, consequently, on the farm location (McBride and Mathews, 2011; USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service, 2017). For instance, a study conducted by USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service, 2017 reported higher percentages of calves born during the spring in the Central US (78.4% of the animals) and West (64.0% of the animals) regions compared to calves born in the East region (43.0%). Additionally, the same study reported that over one-fourth of calves (27.9%) were born during the fall in the East region of the US, while only 15.3% and 7.8% were born in the West and Central regions, respectively (USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service, 2017). In the same report, a small proportion of cows calved during the

summer [2.7% (USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service, 2017)]. Ultimately, the pattern of calves born per season, as presented by the USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service, 2017, is in agreement with our findings (10% summer, 18% fall, 30% spring, and 42% winter; **Supplementary Table S1**). Therefore, the season might reflect the region where most calves were born. In other words, animals born during the fall, winter, and spring mainly represent animals from the East, North, and Central-West regions, respectively. Another theory has to do with the size of the herd and/or contemporary group. Animals born during low season probably have few lot-mates and, as a consequence, lower probability of stressful social interactions.

Animal behavior has been associated with memory ability, in which positive and frequent human–animal interaction could positively impact future behavioral responses (Boivin et al., 2009). The creep feeding system and additional phenotypic measurement (e.g., ultrasound information) are generally associated with additional interaction with humans (Broadhead et al., 2018). A temperament study in Angus cows in Bulgaria reported that more docile cows have frequent contact with people (Karamfilov, 2022). Our phenotypic analyses showed a positive relationship between the use of creep feeding and if the animal had additional measurements (i.e., ultrasound and feed intake). On the other hand, an experimental study reported that despite efforts to give positive animal–human experiences during the handling, cattle would still be averse to the handling process, and its fear would increase (Petherick et al., 2002).

These findings provide a framework for understanding how sex and extrinsic factors influence temperament at the phenotypic level. Furthermore, it can provide insights to improve farm management to guide future epigenetic experimental studies. Environmental conditions regulating gene expression are well known for many traits, including docility (Cantrell et al., 2019). Hence, the listed environmental factors affecting yearling temperament can also be indicative of a programmer of epigenetics modifications.

Model Choice and Parameter Estimation

The ultimate goal of model definition is to facilitate genetic improvement through more accurate estimation of variance components and breeding values (Bijma, 2006). Therefore, four animal models were tested in this study, including direct genetic and maternal effects (**Tables 2**, **4**). Genetic parameters for the DMGP model, which includes direct genetic, maternal genetic, and maternal environmental effects, did not converge, which might be due to inherently problematic estimation of maternal effects (Meyer, 1994). The small average number of progenies per dam and grand-dam in our dataset represents an additional challenge for the estimation of maternal effects [i.e., average of 1.6 ± 1.82 progenies per dams (484,322 dams in the pedigree), 2.2 ± 4.9 grandchildren per grand-dam (total of 338,628 grand-dams in the pedigree), and 2.9 ± 10.2 great-grandchildren per great-grand-dam (total of 250,119 great-grand-dams in the pedigree)].

Among the converged models, the DMG model (i.e., including direct and maternal genetic effects) provided a slightly lower residual variance, suggesting the components are better capturing the total variance. Beckman et al. (2007) have also observed a positive effect of maternal effects (i.e., genetic and environment) on temperament in Limousin cattle. On the other hand, the other models presented slightly desirable dispersion and accuracy. Nevertheless, small differences were observed among the models considering the genetic parameters and linear regression metrics for the best fit model. Thereupon, in routine genetic evaluations, we suggest the use of the D model due to its computational efficiency (time of analysis required), and it was as suitable as the other models. However, as mentioned before, the current data structure might be suboptimal for maternal effect estimation. Therefore, when more data are available, the re-estimation of maternal effects for docility is recommended.

Medium-to-high heritability are in close agreement with previous heritability estimates for cattle temperament using pedigree (Burrow, 2001; Beckman et al., 2007; Haskell et al., 2014) or genomic information (Costilla et al., 2020). Therefore, yearling temperament measured on a 1–6 scale presents sufficient variability to respond to genetic selection, even though a small genetic improvement has been observed across the years in Angus cattle (**Supplementary Figure S3**). Finally, the maternal genetic effects contributed to 4% of the total variation in yearling temperament on the liability scale. Previous studies have also reported low maternal genetic contribution on temperament variation (Burrow, 2001; Beckman et al., 2007). Additionally, a strong negative genetic correlation was observed between additive direct and maternal genetic effects (−0.40 ± 0.039; **Table 4**), which can be explained by the environmental covariances between dam and offspring records and on the fixed effects' structure used in the statistical models for the data analyses (Bijma, 2006). Additionally, the effect of sire-by-herd interaction was not included, which could potentially adjust the negative correlation (Willham, 2000). Negative genetic correlation estimates between direct and maternal additive genetic effects for other traits were also reported in the literature (Dodenhoff et al., 1999; Burrow, 2001; Bijma, 2006; Beckman et al., 2007). However, Willham (2000) stated that there is a possibility of negative correlation between maternal and direct genetic components. The aggressiveness of the dams during the nursery event could be associated with higher protectiveness of the progeny and not necessarily with overall temperament of the cow.

Approximated Genetic Correlation Between Yearling Temperament and Economically Relevant and Indicator Traits

In general, Pearson correlation between EBVs (or EPDs) does not properly represent the genetic correlation, especially between lowly accurate values (Calo et al., 1973). Therefore, Calo et al. (1973) proposed an approach to adjust for the EBV accuracies, which has been extensively used in dairy cattle studies (Dikmen et al., 2012; Costa et al., 2020; Dechow et al., 2020). Traditionally,

dairy cattle have higher EBV accuracies than beef cattle populations. Thereof, the method used by Calo et al. (1973) might penalize Pearson correlations when correlated traits have low accuracies. For this study, we calculated a correlation of EBVs weighted for the accuracies.

Pairwise genetic correlation among all other relevant traits had been performed, and similar direction and scale were observed between genetic correlation using the weighted Pearson correlation and multi-trait animal model, considering the same population (www.angus.org/Nce/Heritabilities). For completeness, we have presented the approximated genetic correlations between all pairs of traits. However, only associations between yearling temperament and the other traits are discussed in this article.

Positive and favorable genetic correlation was observed between temperament and growth, feed efficiency, precocity, and carcass traits. Similar results were observed in cattle. For instance, temperament was genetically associated with higher average daily gain, better conformation scores, finishing precocity, and muscling (Burrow, 2001; King et al., 2006; Phocas et al., 2006; Hoppe et al., 2010; Sant'Anna et al., 2012, 2019; Cooke, 2014; Coutinho et al., 2017). Nervous cattle tend to allocate more energy into the state of excitement instead of using it for other physiological functions such as growth and reproduction (Petherick et al., 2002). In terms of carcass quality, the study observed alterations in the carcass pH based on the temperament groups [i.e., more docile vs. more aggressive; (Petherick et al., 2002)]. The pH variations by animal temperament were speculated to be caused by differences in lactic acid concentrations, as nervous animals have higher levels of lactic acid as a consequence of the distress (Petherick et al., 2002). Other traits (e.g., foot score and hair shedding) had low genetic correlation with temperament, suggesting its partially independent genetic responses.

In summary, there were no unfavorable correlations between temperament and other relevant traits in livestock. This suggests that direct selection for temperament would not negatively impact the genetic improvement of other relevant traits. Actually, the long-term selection for growth, feed efficiency, precocity, and carcass traits has favorable effect on the indirect selection for temperament, or *vice versa*.

Weighted Single-Step Genome-Wide Association Analyses

Animal temperament is a behavioral response from a multifactorial process. The starting point is a stimulus, which could be from a simple touch, sight, sound, smell, or taste, until more complex levels, for instance, an intuitive feeling of dominance. The complexity extends to the genetic level, in which behavior is controlled by many genes with small effect (Costilla et al., 2020; Alvarenga et al., 2021). Despite the high heritability estimate, our study found three pieces of evidence of a polygenic nature of yearling temperament in North American Angus. First, an even distribution of variance explained by genomic windows was observed across the genome (**Figure 3**). Second, none of the genomic windows explained the large proportion of the total additive genetic variance for yearling temperament—the maximum variance explained was 0.51% on BTA11 (**Supplementary Table S4**). Finally, at the modeling level, small differences were observed in the distribution of the genomic window effect and variance explained when using ssGWAS or WssGWAS approaches (**Supplementary Figure S2**). Similar findings were observed by Araujo et al. (2022) while fitting haplotype groups instead of SNP effects using the AAA dataset for temperament. In general, a weighting method (e.g., WssGWAS) is expected to prioritize major regions when a less polygenic trait is being analyzed (Wang et al., 2014). In summary, the weighting process for genomic associations using SNPs is not necessary if the target trait underlies a polygenic architecture.

At the genomic level, another characteristic of yearling temperament is the influence of the X-chromosome. In comparison with other chromosomes, X was underrepresented with a total of 97 SNPs (59.13 ± 45.90 Kbp is the average ± standard deviation distance between SNPs) out of 42,662 SNPs (58.18 ± 54.47), because we kept only pseudo-autosomal SNPs that probably follow a similar recombination pattern to that of the autosomal chromosomes. Regardless, a peak was captured on the X chromosome, but no annotated genes were identified within this region. However, this finding highlights the importance of the X-chromosome for behavioral traits, as reported by Alvarenga et al. (2021). In this context, Musante and Ropers (2014) have reviewed numerous X-chromosome defects linked to cognitive disorders in humans. Therefore, we encouraged further studies evaluating the impact of the X-chromosome on the target trait, including both PAR and non-PAR regions. Song et al. (2021) have evaluated few alternative approaches to account for the sex-linked chromosomes in genomic analyses, such as an interaction of sex-linked SNP and sex.

The QTL annotation within the selected genomic regions corroborates and potentially justifies the genetic correlation between temperament and growth, carcass traits, precocity, reproduction, and maternal traits (**Figure 2**). The overlapping regions offer insights into mechanisms that can cause the genetic correlation, such as pleiotropy, LD between trait-specific QTLs, and/or LD between QTL and marker (Gianola et al., 2015). The majority of the QTLs were annotated for milk content and yield, which goes in line with the extended list of association studies for milk traits and annotated in the animal QTL database. No straightforward relationship between behavioral traits and milk yield have been drawn (Hedlund and Lovlie, 2015). However, many studies observed that nervous cows tend to have lower milk production, justified by differential hormonal functions (Hedlund and Lovlie, 2015). One gene annotated for milking speed (i.e., *TOX*) is involved in the immune system (Aliahmad et al., 2012). So far, to the best of our knowledge, no direct connection has been made between behavioral traits and the *TOX* gene. In human studies, there are reports indicating that the immune system can be a mediator of adulthood behavior, for example, dysregulation of immune responses was associated with higher aggression (Takahashi et al., 2018). In cattle, *TOX* was reported in association with puberty in Brahman beef cattle (Fortes et al., 2012) and carcass traits in Korean Hanwoo cattle (Bhuiyan et al., 2018), and it has been identified as a

signature of selection in tropical adapted crossbred cattle (Cheruiyot et al., 2018).

In beef cattle, yearling temperament is associated with the length of productive life of the animal (Oliveira et al., 2020), as it is one of the voluntary culling reasons. QTLs annotated for longevity also overlapped with regions for yearling temperament. The genes located in regions for longevity and yearling temperament are mainly associated and enriched in visual mechanisms (Chang et al., 2009). The majority of the genes are from the crystalline gamma family genes (e.g., *CRYGB*, *CRYGC*, and *CRYGD*). In behavior, visual perception is required for the reception of a stimulus, which is subsequently converted to a physiological messenger, and properly recognized as a signal (Chang et al., 2009).

Another term identified in this study which has no significance (p = 0.42) but is biologically important is methylation (**Supplementary Table S5**). Altered cattle behavior has been linked with differential methylation of DNA (Corvo et al., 2020; Littlejohn et al., 2020). Three genes were involved in this methylation process (i.e., *PDE6C*, *RBP4*, and *CRYGC*), and out of those, *PDE6C* was also enriched in the visual perception. In animal models (i.e., mice), this gene has been reported in retinol progenitor regions to be highly methylated, which could impact photoreceptor mechanisms and, consequently, the signaling cascades (Dvoriantchikova et al., 2019). In summary, this study identified genes participating in the visual mechanism, in which the eyes, in addition to being responsible for stimulus reception, have been defined as the window to the brain. For instance, pupil dilation has been associated with cognitive ability and Alzheimer's disease (Granholm et al., 2017; Ozeri-Rotstain et al., 2020).

Another gene participating in the methylation process is *RBP4* (retinol binding protein gene), in which knockout mice for it showed impairment, reduced activity, and increased anxiety-like behavior (Buxbaum et al., 2014). The protein produced by the *RBP4* gene is the principal retinol (i.e., vitamin A) serum transporter in the human body (Reay and Cairns, 2020). Disruption of the transport, metabolism, and signaling of vitamin A metabolites have been associated with mental and behavioral disorders in humans [reviewed by Reay and Cairns (2020)]. In cattle, *RBP4* has been associated with growth traits, which might be related to insulin metabolism (Wang et al., 2010) and heifer fertility (Abdollahi-Arpanahi et al., 2019).

The region with the greatest additive genetic variance explained is located on BTA11, but no annotated gene was found in this region. However, there was one long noncoding RNA (lncRNA, ENSBTAG00000048646) located in a region explaining 0.49% of the total variation (**Supplementary Table S4**). Some lncRNA genes are specifically expressed in brain cells, which implies its roles in neuronal development and cognitive and behavioral regulation (Zhang et al., 2020). In cattle, this gene biotype has been associated with possible skin pigmentation (Weikard et al., 2013), meat quality (Billerey et al., 2014), and lipid metabolism and has been suggested as a possible heat-stress biomarker (Ibeagha-Awemu and Zhao, 2015). However, to the best of our knowledge, no study evaluated the impact of lncRNA on behavior in cattle.

Insulin metabolism is another pathway in which few of the genes identified in this study play an important role, for example, omega-3 Fatty Acid Receptor-1 gene (*FFAR4*) located on BTA26 (Marcotte et al., 2017). Insulin metabolism dysfunction could impact performance and affect the overall mood and cognitive ability of the individuals through dopaminergic abnormal functioning, as observed in humans (Kleinridders et al., 2015; Lyrae Silva et al., 2019). Similar to other genes discussed in this study (e.g., *TOX*), *FFAR4* receptors have also been reported as triggers to inflammatory responses (Oh et al., 2010). In general, *FFAR4* gene signals either a pathway involved in $G_{\alpha q}$, which is linked with insulin resistance with action at the adipocyte, or β-arretin-2, which directly influences the inflammatory pathways with action at the macrophage [the complete process was reviewed and summarized by Oh and Walenta (2014)].

Finally, Alvarenga et al. (2021) systematically reviewed genes associated with farmed mammals' behavior. Among all genes catalogued by the authors, two (i.e., *U6* and *CEP55*) genes previously reported to be associated with behavioral traits in livestock overlapped with the genes identified in this study (Alvarenga et al., 2021). Interestingly, the U6 spliceosomal RNA (*U6*) gene was previously reported to be associated with temperament (Hanna et al., 2014), maternal behavior (Michenet et al., 2016), and sucking reflex (Dreher et al., 2019) in cattle and feeding behavior (Cross et al., 2018) and adrenaline/creatine level (Terenina et al., 2013) in pigs. The *U6* gene is a noncoding RNA gene; in this study, the paralog gene identified is on BTA8 explaining 0.22% of the total genetics. The U6 snRNA is highly conserved in eukaryotes, and it is located in the spliceosome, where it orchestrates the splicing function (Didychuk et al., 2018), which seems to have an evolutionary importance in terms of organism viability. Molecular and chemistry studies might be crucial to underpin the role of uridine-rich small nuclear RNAs (i.e., *U6*) in alternative animal behavior. However, more studies are required to elucidate the roles of the genes identified in this study, including *in vitro* gene knockout and gene editing studies.

CONCLUSIONS

Our study encompasses a large, diverse, and robust dataset from the North American Angus cattle population. Yearling temperament was shown to be heritable (0.38 ± 0.01), suggesting that genetic gains can be effectively obtained through direct genetic and genomic selection. Additionally, direct selection for yearling temperament is not expected to result in unfavorable effects on other economically important traits. Actually, the long-term selection for growth, feed efficiency, precocity, and carcass traits has a favorable effect on the indirect selection for temperament, or *vice versa*.

Sex and extrinsic factors affect animal behavior, such as age of dam, additional human–animal interaction, and birth region. Fitting maternal effects are not crucial for breeding value estimation of temperament in the US Angus population,

suggesting the use of the reduced model due to its computational efficiency. Many of the genomic regions associated with yearling temperament were enriched in QTL regions linked with milking speed, longevity, and other economically important traits. Pseudo-autosomal regions of the X-chromosome seem to play a role in yearling temperament, as well as RNA gene biotypes, for example, long noncoding RNA genes. Finally, many of the genes were enriched in visual pathways, which in addition to being important for stimulus, have also been linked to cognitive abilities. The SNPs and genomic regions identified in this study can be used when designing customized SNP panels, to be further investigated in functional studies aiming to better understand the biological mechanisms influencing cattle temperament, as well as used for a biology-driven genomic prediction. In summary, the insights from our study will be useful for various practical applications and future research, including the breeding industry, farm management, and behavioral genomics studies.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: The data supporting the results of this article are included within the article and in its **Supplementary Files**. The raw data cannot be made available, as it is property of the American Angus Association (Saint Joseph, MO, United States) and this information is commercially sensitive. Requests to access these datasets should be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AA partially conceptualized the article idea, analyzed the data, interpreted and discussed the results, and prepared the original draft, including writing and reviewing; HO provided technical support for the analysis, reviewed, and edited. SM participated in the conceptualization and reviewed the final version. FS provided technical support for the analysis and reviewed the final manuscript. LB conceptualized and supervised the project, provided training to the first author, and reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This project was partially funded by the American Angus Association and Angus Genetics Inc. (MO, United States).

ACKNOWLEDGMENTS

The authors acknowledge American Angus Association and Angus Genetics Inc. staff for all the support in regard to data availability and computing resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.794625/full#supplementary-material

REFERENCES

Abdollahi-Arpanahi, R., Carvalho, M. R., Ribeiro, E. S., and Peñagaricano, F. (2019). Association of Lipid-Related Genes Implicated in Conceptus Elongation with Female Fertility Traits in Dairy Cattle. *J. Dairy Sci.* 102, 10020–10029. doi:10.3168/jds.2019-17068

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, 743–752. doi:10.3168/jds.2009-2730

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., and Wang, H. (2014). "PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs," in Abstract retrieved from 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, August 17–22, 2014. doi:10.13140/2.1.4801.5045

Aliahmad, P., Seksenyan, A., and Kaye, J. (2012). The many Roles of TOX in the Immune System. *Curr. Opin. Immunol.* 24, 173–177. doi:10.1016/j.coi.2011.12.001

Altman, D. G., and Bland, J. M. (2005). Standard Deviations and Standard Errors. *BMJ* 331, 903. doi:10.1136/bmj.331.7521.903

Alvarenga, A. B., Oliveira, H. R., Chen, S.-Y., Miller, S. P., Marchant-Forde, J. N., Grigoletto, L., et al. (2021). A Systematic Review of Genomic Regions and Candidate Genes Underlying Behavioral Traits in Farmed Mammals and Their Link with Human Disorders. *Animals* 11, 715. doi:10.3390/ani11030715

Araujo, A. C., Carneiro, P. L. S., Alvarenga, A. B., Oliveira, H. R., Miller, S. P., Brito, L. B., et al. (2022). Haplotype-based Single-step GWAS for Yearling Temperament in American Angus Cattle. *Genes* 13, 17. doi:10.3390/genes13010017

Barrozo, D., Buzanskas, M. E., Oliveira, J. A., Munari, D. P., Neves, H. H. R., and Queiroz, S. A. (2012). Genetic Parameters and Environmental Effects on Temperament Score and Reproductive Traits of Nellore Cattle. *Animal* 6, 36–40. doi:10.1017/S1751731111001169

Beckman, D. W., Enns, R. M., Speidel, S. E., Brigham, B. W., and Garrick, D. J. (2007). Maternal Effects on Docility in Limousin Cattle1. *J. Anim. Sci.* 85, 650–657. doi:10.2527/jas.2006-450

Bhuiyan, M. S. A., Lim, D., Park, M., Lee, S., Kim, Y., Gondro, C., et al. (2018). Functional Partitioning of Genomic Variance and Genome-wide Association Study for Carcass Traits in Korean Hanwoo Cattle Using Imputed Sequence Level SNP Data. *Front. Genet.* 9, 1–14. doi:10.3389/fgene.2018.00217

Bijma, P. (2006). Estimating Maternal Genetic Effects in Livestock1. *J. Anim. Sci.* 84, 800–806. doi:10.2527/2006.844800x

Billerey, C., Boussaha, M., Esquerré, D., Rebours, E., Djari, A., Meersseman, C., et al. (2014). Identification of Large Intergenic Non-coding RNAs in Bovine Muscle Using Next-Generation Transcriptomic Sequencing. *BMC Genomics* 15, 1. doi:10.1186/1471-2164-15-499

Blake, J. A., Dolan, M., Drabkin, H., Hill, D. P., Ni, L., Sitnikov, D., et al. (2013). Gene Ontology Annotations and Resources. *Nucleic Acids Res.* 41, D530–D535. doi:10.1093/nar/gks1050

Boivin, X., Gilard, F., and Egal, D. (2009). The Effect of Early Human Contact and the Separation Method from the Dam on Responses of Beef Calves to Humans. *Appl. Anim. Behav. Sci.* 120, 132–139. doi:10.1016/j.applanim.2009.05.011

Broadhead, D. L., Stockton, M., Stalker, L. A., Musgrave, J. A., and Funston, R. N. (2018). 418 Economics of Creep Feeding a Spring Calving Beef Herd in the Nebraska Sandhills. *J. Anim. Sci.* 96, 224–225. doi:10.1093/jas/sky073.415

Burdick, N. C., Randel, R. D., Carroll, J. A., and Welsh, T. H. (2011). Interactions between Temperament, Stress, and Immune Function in Cattle. *Int. J. Zoolog.* 2011, 1–9. doi:10.1155/2011/373197

Burrow, H. M. (2001). Variances and Covariances between Productive and Adaptive Traits and Temperament in a Composite Breed of Tropical Beef Cattle. *Livestock Prod. Sci.* 70, 213–233. doi:10.1016/S0301-6226(01)00178-6

Burrow, H., Seifert, G., and Corbet, N. (1988). A New Technique for Measuring Temperament in Cattle. *Soc. Anim. Prod.* 17, 54.

Busby, D. (2015). "Lessons Learned from 32 Years of Retained Ownership – TCSCF Summary," in Abstract retrieved from Driftless Region Beef Conference, Iowa, January 22–23, 2015, 29–31.

Buxbaum, J. N., Roberts, A. J., Adame, A., and Masliah, E. (2014). Silencing of Murine Transthyretin and Retinol Binding Protein Genes Has Distinct and Shared Behavioral and Neuropathologic Effects. *Neuroscience* 275, 352–364. doi:10.1016/j.neuroscience.2014.06.019

Calo, L. L., McDowell, R. E., VanVleck, L. D., and Miller, P. D. (1973). Genetic Aspects of Beef Production Among Holstein-Friesians Pedigree Selected for Milk Production1. *J. Anim. Sci.* 37, 676–682. doi:10.2527/jas1973.373676x

Cantrell, B., Lachance, H., Murdoch, B., Sjoquist, J., Funston, R., Weaber, R., et al. (2019). Global DNA Methylation in the Limbic System of Cattle. *Epigenomes* 3, 8. doi:10.3390/epigenomes3020008

Carlström, C., Strandberg, E., Johansson, K., Pettersson, G., Stålhammar, H., and Philipsson, J. (2014). Genetic Evaluation of In-Line Recorded Milkability from Milking Parlors and Automatic Milking Systems. *J. Dairy Sci.* 97, 497–506. doi:10.3168/jds.2013-6948

Chan, F. Y. (2012). "Genome Wide Association Studies for Temperament in New Zealand Dairy Cattle," (Palmerston North, NZ: Massey University). Master's thesis.

Chang, B., Grau, T., Dangel, S., Hurd, R., Jurklies, B., Sener, E. C., et al. (2009). A Homologous Genetic Basis of the Murine Cpfl1 Mutant and Human Achromatopsia Linked to Mutations in the PDE6C Gene. *Proc. Natl. Acad. Sci.* 106, 19581–19586. doi:10.1073/pnas.0907720106

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., and Mujibi, F. D. N. (2018). Signatures of Selection in Admixed Dairy Cattle in Tanzania. *Front. Genet.* 9, 1–15. doi:10.3389/fgene.2018.00607

Christensen, O. F., and Lund, M. S. (2010). Genomic Prediction when Some Animals Are Not Genotyped. *Genet. Sel Evol.* 42, 8. doi:10.1186/1297-9686-42-2

Cole, J. B., VanRaden, P. M., O'Connell, J. R., van Tassell, C. P., Sonstegard, T. S., Schnabel, R. D., et al. (2009). Distribution and Location of Genetic Effects for Dairy Traits. *J. Dairy Sci.* 92, 2931–2946. doi:10.3168/jds.2008-1762

Cooke, R. F. (2014). Bill E. Kunkle Interdisciplinary Beef Symposium: Temperament and Acclimation to Human Handling Influence Growth, Health, and Reproductive Responses in Bos taurus and Bos indicus Cattle1. *J. Anim. Sci.* 92, 5325–5333. doi:10.2527/jas.2014-8017

Cooke, R. F., Bohnert, D. W., Meneghetti, M., Losi, T. C., and Vasconcelos, J. L. M. (2011). Effects of Temperament on Pregnancy Rates to Fixed-Timed AI in Bos indicus Beef Cows. *Livestock Sci.* 142, 108–113. doi:10.1016/j.livsci.2011.06.024

Costa, A., Bovenhuis, H., and Penasa, M. (2020). Changes in Milk Lactose Content as Indicators for Longevity and Udder Health in Holstein Cows. *J. Dairy Sci.* 103, 11574–11584. doi:10.3168/jds.2020-18615

Costilla, R., Kemper, K. E., Byrne, E. M., Porto-Neto, L. R., Carvalheiro, R., Purfield, D. C., et al. (2020). Genetic Control of Temperament Traits across Species: Association of Autism Spectrum Disorder Risk Genes with Cattle Temperament. *Genet. Sel Evol.* 52, 1–14. doi:10.1186/s12711-020-00569-z

Coutinho, M. A. d. S., Ramos, P. M., da Luz e Silva, S., Martello, L. S., Pereira, A. S. C., and Delgado, E. F. (2017). Divergent Temperaments Are Associated with Beef Tenderness and the Inhibitory Activity of Calpastatin. *Meat Sci.* 134, 61–67. doi:10.1016/j.meatsci.2017.06.017

Cross, A. J., Keel, B. N., Brown-Brandl, T. M., Cassady, J. P., and Rohrer, G. A. (2018). Genome-wide Association of Changes in Swine Feeding Behaviour Due to Heat Stress. *Genet. Sel Evol.* 50, 1–12. doi:10.1186/s12711-018-0382-1

Cundiff, L. V., Van Vleck, L. D., and Hohenboken, W. D. (2018). Guidelines for Uniform Beef Improvement Programs (Beef Improvement Federation) Guidelines. Available at: beefimprovement.org/index.php/Guidelines_for_Uniform_Beef_Improvement_Programs (Accessed September, 2021).

Cziszter, L. T., Gavojdian, D., Neamt, R., Neciu, F., Kusza, S., and Ilie, D.-E. (2016). Effects of Temperament on Production and Reproductive Performances in Simmental Dual-Purpose Cows. *J. Vet. Behav.* 15, 50–55. doi:10.1016/j.jveb.2016.08.070

Dechow, C. D., Sondericker, K. S., Enab, A. A., and Hardie, L. C. (2020). Genetic, Farm, and Lactation Effects on Behavior and Performance of US Holsteins in Automated Milking Systems. *J. Dairy Sci.* 103, 11503–11514. doi:10.3168/jds.2020-18786

Del Corvo, M., Bongiorni, S., Stefanon, B., Sgorlon, S., Valentini, A., Ajmone Marsan, P., et al. (2020). Genome-wide DNA Methylation and Gene Expression Profiles in Cows Subjected to Different Stress Level as Assessed by Cortisol in Milk. *Genes* 11, 850–919. doi:10.3390/genes11080850

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, P3. doi:10.1186/gb-2003-4-5-p3

Dickson, D. P., Barr, G. R., Johnson, L. P., and Wieckert, D. A. (1970). Social Dominance and Temperament of Holstein Cows. *J. Dairy Sci.* 53, 904–907. doi:10.3168/jds.S0022-0302(70)86316-0

Didychuk, A. L., Butcher, S. E., and Brow, D. A. (2018). The Life of U6 Small Nuclear RNA, from Cradle to Grave. *RNA* 24, 437–460. doi:10.1261/rna.065136.117

Dikmen, S., Cole, J. B., Null, D. J., and Hansen, P. J. (2012). Heritability of Rectal Temperature and Genetic Correlations with Production and Reproduction Traits in Dairy Cattle. *J. Dairy Sci.* 95, 3401–3405. doi:10.3168/jds.2011-4306

Dodenhoff, J., Van Vleck, L. D., and Wilson, D. E. (1999). Comparison of Models to Estimate Genetic Effects of Weaning Weight of Angus Cattle. *J. Anim. Sci.* 77, 3176–3184. doi:10.2527/1999.77123176x

Dos Santos, F. C., Peixoto, M. G. C. D., Fonseca, P. A. d. S., Pires, M. d. F. Á., Ventura, R. V., Rosse, I. d. C., et al. (2017). Identification of Candidate Genes for Reactivity in Guzerat (Bos indicus) Cattle: A Genome-wide Association Study. *PLoS ONE* 12, e0169163–15. doi:10.1371/journal.pone.0169163

Dreher, C., Wellmann, R., Stratz, P., Schmid, M., Preuß, S., Hamann, H., et al. (2019). Genomic Analysis of Perinatal Sucking Reflex in German Brown Swiss Calves. *J. Dairy Sci.* 102, 6296–6305. doi:10.3168/jds.2019-16487

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package biomaRt. *Nat. Protoc.* 4, 1184–1191. doi:10.1038/jid.2014.371

Dvoriantchikova, G., Seemungal, R. J., and Ivanov, D. (2019). DNA Methylation Dynamics during the Differentiation of Retinal Progenitor Cells into Retinal Neurons Reveal a Role for the DNA Demethylation Pathway. *Front. Mol. Neurosci.* 12, 1–9. doi:10.3389/fnmol.2019.00182

Ede, T., Lecorps, B., von Keyserlingk, M. A. G., and Weary, D. M. (2019). Calf Aversion to Hot-Iron Disbudding. *Sci. Rep.* 9, 1–6. doi:10.1038/s41598-019-41798-7

Esmailizadeh, A. K., Bottema, C. D. K., Sellick, G. S., Verbyla, A. P., Morris, C. A., Cullen, N. G., et al. (2008). Effects of the Myostatin F94L Substitution on Beef Traits1. *J. Anim. Sci.* 86, 1038–1046. doi:10.2527/jas.2007-0589

Fortes, M. R. S., Lehnert, S. A., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N. J., et al. (2012). Finding Genes for Economically Important Traits: Brahman Cattle Puberty. *Anim. Prod. Sci.* 52, 143–150. doi:10.1071/AN11165

Garza-Brenner, E., Sifuentes-Rincón, A. M., Randel, R. D., Paredes-Sánchez, F. A., Parra-Bracamonte, G. M., Arellano Vera, W., et al. (2017). Association of SNPs in Dopamine and Serotonin Pathway Genes and Their Interacting Genes with Temperament Traits in Charolais Cows. *J. Appl. Genet.* 58, 363–371. doi:10.1007/s13353-016-0383-0

Gauly, M., Mathiak, H., Hoffmann, K., Kraus, M., and Erhardt, G. (2001). Estimating Genetic Variability in Temperamental Traits in German Angus and Simmental Cattle. *Appl. Anim. Behav. Sci.* 74, 109–119. doi:10.1016/S0168-1591(01)00151-4

Geweke, J. (1991). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Bayesian Statistics*. Minneapolis, USA: Federal Reserve Bank of Minneapolis, Research Department, 169–193.

Gianola, D., de los Campos, G., Toro, M. A., Naya, H., Schön, C.-C., and Sorensen, D. (2015). Do molecular Markers Inform about Pleiotropy? *Genetics* 201, 23–29. doi:10.1534/genetics.115.179978

Glenske, K., Prinzenberg, E.-M., Brandt, H., Gauly, M., and Erhardt, G. (2011). A Chromosome-wide QTL Study on BTA29 Affecting Temperament Traits in German Angus Beef Cattle and Mapping of DRD4. *Animal* 5, 195–197. doi:10.1017/S1751731110001801

Granholm, E. L., Panizzon, M. S., Elman, J. A., Jak, A. J., Hauger, R. L., Bondi, M. W., et al. (2017). Pupillary Responses as a Biomarker of Early Risk for Alzheimer's Disease. *Jad* 56, 1419–1428. doi:10.3233/JAD-161078

Greenwood, P. L. (2021). Review: An Overview of Beef Production from Pasture and Feedlot Globally, as Demand for Beef and the Need for Sustainable Practices Increase. *Animal* 15, 100295. doi:10.1016/j.animal.2021.100295

Haskell, M. J., Simm, G., and Turner, S. P. (2014). Genetic Selection for Temperament Traits in Dairy and Beef Cattle. *Front. Genet.* 5, 1–18. doi:10.3389/fgene.2014.00368

Hearnshaw, H., and Morris, C. (1984). Genetic and Environmental Effects on a Temperament Score in Beef Cattle. *Aust. J. Agric. Res.* 35, 723–733. doi:10.1071/AR9840723

Hedlund, L., and Løvlie, H. (2015). Personality and Production: Nervous Cows Produce Less Milk. *J. Dairy Sci.* 98, 5819–5828. doi:10.3168/jds.2014-8667

Heidelberger, P., and Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Res.* 31, 1109–1144. doi:10.1287/opre.31.6.1109

Hidalgo, J., Tsuruta, S., Lourenco, D., Masuda, Y., Huang, Y., Gray, K. A., et al. (2020). Changes in Genetic Parameters for Fitness and Growth Traits in Pigs under Genomic Selection. *J. Anim. Sci.* 98, 1–12. doi:10.1093/jas/skaa032

Hoppe, S., Brandt, H. R., König, S., Erhardt, G., and Gauly, M. (2010). Temperament Traits of Beef Calves Measured under Field Conditions and Their Relationships to Performance1. *J. Anim. Sci.* 88, 1982–1989. doi:10.2527/jas.2008-1557

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

Hulsman Hanna, L. L., Garrick, D. J., Gill, C. A., Herring, A. D., Riggs, P. K., Miller, R. K., et al. (2014). Genome-wide Association Study of Temperament and Tenderness Using Different Bayesian Approaches in a Nellore-Angus Crossbred Population. *Livestock Sci.* 161, 17–27. doi:10.1016/j.livsci.2013.12.012

Ibeagha-Awemu, E. M., and Zhao, X. (2015). Epigenetic marks: Regulators of Livestock Phenotypes and Conceivable Sources of Missing Variation in Livestock Improvement Programs. *Front. Genet.* 6, 1–17. doi:10.3389/fgene.2015.00302

Johnson, T., Keehan, M., Harland, C., Lopdell, T., Spelman, R. J., Davis, S. R., et al. (2019). Short Communication: Identification of the Pseudoautosomal Region in the Hereford Bovine Reference Genome Assembly ARS-UCD1.2. *J. Dairy Sci.* 102, 3254–3258. doi:10.3168/jds.2018-15638

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Karamfilov, S. (2022). Study on the Temperament of Cows of the Aberdeen Angus Cattle Breed. *Czech J. Anim. Sci.* 67, 8–14. doi:10.17221/88/2021-cjas

Kasimanickam, V., Abdel Aziz, R., Williams, H., and Kasimanickam, R. (2018). Predictors of Beef Calf Temperament at Weaning and its Impact on Temperament at Breeding and Reproductive Performance. *Reprod. Dom Anim.* 53, 484–494. doi:10.1111/rda.13135

Kilgour, R. (1975). The Open-Field Test as an Assessment of the Temperament of Dairy Cows. *Anim. Behav.* 23, 615–624. doi:10.1016/0003-3472(75)90139-6

King, D. A., Schuehle Pfeiffer, C. E., Randel, R. D., Welsh, T. H., Oliphint, R. A., Baird, B. E., et al. (2006). Influence of Animal Temperament and Stress Responsiveness on the Carcass Quality and Beef Tenderness of Feedlot Cattle. *Meat Sci.* 74, 546–556. doi:10.1016/j.meatsci.2006.05.004

Kleinridders, A., Cai, W., Cappellucci, L., Ghazarian, A., Collins, W. R., Vienberg, S. G., et al. (2015). Insulin Resistance in Brain Alters Dopamine Turnover and Causes Behavioral Disorders. *Proc. Natl. Acad. Sci. USA* 112, 3463–3468. doi:10.1073/pnas.1500877112

Legarra, A., and Reverter, A. (2018). Semi-parametric Estimates of Population Accuracy and Bias of Predictions of Breeding Values and Future Phenotypes Using the LR Method. *Genet. Sel Evol.* 50, 1–18. doi:10.1186/s12711-018-0426-6

Lenth, R. V. (2016). Least-Squares Means: TheRPackagelsmeans. *J. Stat. Soft.* 69. doi:10.18637/jss.v069.i01

Littlejohn, B. P., Price, D. M., Neuendorff, D. A., Carroll, J. A., Vann, R. C., Riggs, P. K., et al. (2020). Influence of Prenatal Transportation Stress-Induced Differential DNA Methylation on the Physiological Control of Behavior and Stress Response in Suckling Brahman Bull Calves. *J. Anim. Sci.* 98, 1–18. doi:10.1093/jas/skz368

Lu, D., Sargolzaei, M., Kelly, M., Li, C., Vander Voort, G., Wang, Z., et al. (2012). Linkage Disequilibrium in Angus, Charolais, and Crossbred Beef Cattle. *Front. Gene* 3, 1–10. doi:10.3389/fgene.2012.00152

Lyrae Silva, N. D. M., Lam, M. P., Soares, C. N., Munoz, D. P., Milev, R., and De Felice, F. G. (2019). Insulin Resistance as a Shared Pathogenic Mechanism between Depression and Type 2 Diabetes. *Front. Psychiatry* 10. doi:10.3389/fpsyt.2019.00057

Macleod, I. M., Bowman, P. J., Chamberlain, A. J., Vander Jagt, C. J., Daetwyler, H. D., Hayes, B. J., et al. (2019). "Genomic Prediction and Candidate Gene Discovery for Dairy Cattle Temperament Using Sequence Data and Functional Biology," in Abstract retrieved from the 23rd Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW, October 27–November 1, 2019, 416–419.

McBride, W. D., and Mathews, K. (2011). The Diverse Structure and Organization of U.S. Beef Cow-Calf Farms. Available at: www.ers.usda.gov/webdocs/publications/44530/7611_eib73.pdf?v=0 (Accessed September, 2021).

Medrano, J. F. (2017). "The New Bovine Reference Assembly and its Value for Genomic Research," in Abstract retrieved from the 21st Association for the Advancement of Animal Breeding and Genetics, Queensland, Australia, July 2–5, 2017, 161–166.

Meyer, K. (1994). Estimates of Direct and Maternal Correlations Among Growth Traits in Australian Beef Cattle. *Livestock Prod. Sci.* 38, 91–105. doi:10.1016/0301-6226(94)90053-1

Michenet, A., Saintilan, R., Venot, E., and Phocas, F. (2016). Insights into the Genetic Variation of Maternal Behavior and Suckling Performance of continental Beef Cows. *Genet. Sel Evol.* 48, 1–12. doi:10.1186/s12711-016-0223-z

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing Procedures for Genetic Evaluation Including Phenotypic, Full Pedigree, and Genomic Information. *J. Dairy Sci.* 92, 4648–4655. doi:10.3168/jds.2009-2064

Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., VanRaden, P. M., and Lawlor, T. J. (2013). Methods to Approximate Reliabilities in Single-step Genomic Evaluation. *J. Dairy Sci.* 96, 647–654. doi:10.3168/jds.2012-5656

Musante, L., and Ropers, H. H. (2014). Genetics of Recessive Cognitive Disorders. *Trends Genet.* 30, 32–39. doi:10.1016/j.tig.2013.09.008

Neja, W., Sawa, A., Jankowska, M., Bogucki, M., and Krężel-Czopek, S. (2015). Effect of the Temperament of Dairy Cows on Lifetime Production Efficiency. *Arch. Anim. Breed.* 58, 193–197. doi:10.5194/aab-58-193-2015

Northcutt, S., and Bowman, B. (2007). By the Numbers: Docility Genetic Evaluation Research. Available at: www.angus.org/Nce/Documents/ByThenumbersDocility.pdf (Accessed September, 2021).

Oh, D. Y., Talukdar, S., Bae, E. J., Imamura, T., Morinaga, H., Fan, W., et al. (2010). GPR120 Is an omega-3 Fatty Acid Receptor Mediating Potent Anti-inflammatory and Insulin-Sensitizing Effects. *Cell* 142, 687–698. doi:10.1016/j.cell.2010.07.041

Oh, D. Y., and Walenta, E. (2014). Omega-3 Fatty Acids and FFAR4. *Front. Endocrinol.* 5, 2–6. doi:10.3389/fendo.2014.00115

Oliveira, H. R., Brito, L. F., Miller, S. P., and Schenkel, F. S. (2020). Using Random Regression Models to Genetically Evaluate Functional Longevity Traits in North American Angus Cattle. *Animals* 10, 2410–2430. doi:10.3390/ani10122410

Ozeri-Rotstain, A., Shachaf, I., Farah, R., and Horowitz-Kraus, T. (2020). Relationship between Eye-Movement Patterns, Cognitive Load, and reading Ability in Children with reading Difficulties. *J. Psycholinguist Res.* 49, 491–507. doi:10.1007/s10936-020-09705-8

Petherick, J. C., Holroyd, R. G., Doogan, V. J., and Venus, B. K. (2002). Productivity, Carcass and Meat Quality of Lot-Fed Bos indicus Cross Steers Grouped According to Temperament. *Aust. J. Exp. Agric.* 42, 389–398. doi:10.1071/EA01084

Phocas, F., Boivin, X., Sapa, J., Trillat, G., Boissy, A., and Le Neindre, P. (2006). Genetic Correlations between Temperament and Breeding Traits in Limousin Heifers. *Anim. Sci.* 82, 805–811. doi:10.1017/ASC200696

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reay, W. R., and Cairns, M. J. (2020). The Role of the Retinoids in Schizophrenia: Genomic and Clinical Perspectives. *Mol. Psychiatry* 25, 706–718. doi:10.1038/s41380-019-0566-2

Reinhardt, V., Reinhardt, C., and Reinhardt, A. (1987). Evaluating Sex Differences in Aggressiveness in Cattle, Bison and Rhesus Monkeys. *Behav* 102, 58–66. Available at:

brill.com/view/journals/beh/102/1-2/article-p58_4.xml?casa_token=_Lc1kNXVu_8AAAAA:itN_-I1NB9w7VZiB40fpS0uZ_oCkfajHcIuM1sBBSm9Nur9jV2BeVX1Q-rvl0XpQkIKQh7CTyA. doi:10.1163/156853986x00045 Accessed February 14, 2022).

Riley, D. G., Gill, C. A., Boldt, C. R., Funkhouser, R. R., Herring, A. D., Riggs, P. K., et al. (2016). Crossbred Bos indicus Steer Temperament as Yearlings and Whole Genome Association of Steer Temperament as Yearlings and Calf Temperament post-weaning12. *J. Anim. Sci.* 94, 1408–1414. doi:10.2527/jas.2015-0041

Rolfe, K. M., Snelling, W. M., Nielsen, M. K., Freetly, H. C., Ferrell, C. L., and Jenkins, T. G. (2011). Genetic and Phenotypic Parameter Estimates for Feed Intake and Other Traits in Growing Beef Cattle, and Opportunities for Selection123. *J. Anim. Sci.* 89, 3452–3459. doi:10.2527/jas.2011-3961

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Zimin, A., et al. (2018). "Modernizing the Bovine Reference Genome Assembly," in 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand, February 11–16, 2018, 6–11.

Sant'Anna, A. C., Paranhos da Costa, M. J. R., Baldi, F., Rueda, P. M., and Albuquerque, L. G. (2012). Genetic Associations between Flight Speed and Growth Traits in Nellore Cattle1. *J. Anim. Sci.* 90, 3427–3432. doi:10.2527/jas.2011-5044

Sant'anna, A. C., Valente, T. D. S., Magalhães, A. F. B., Espigolan, R., Ceballos, M. C., De Albuquerque, L. G., et al. (2019). Relationships between Temperament, Meat Quality, and Carcass Traits in Nellore Cattle1. *J. Anim. Sci.* 97, 4721–4731. doi:10.1093/jas/skz324

Santos, B., Archer, J. A., Martin-Collado, D., Quinton, C. D., Crowley, J., Amer, P. R., et al. (2019). "Industry Consultation Survey for the American Angus $Value Indexes Review," in 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW, October 27–November 1, 2019.

Sato, S. (1981). Factors Associated with Temperament of Beef Cattle. *Nihon Chikusan Gakkaiho* 52, 595–605. doi:10.2508/chikusan.52.595

Smith, B. J. (2007). boa: AnRPackage for MCMC Output Convergence Assessment and Posterior Inference. *J. Stat. Soft.* 21. doi:10.18637/jss.v021.i11

Song, Y., Biernacka, J. M., and Winham, S. J. (2021). Testing and Estimation of X-Chromosome SNP Effects: Impact of Model Assumptions. *Genet. Epidemiol.* 45 (6), 577–592. doi:10.1002/gepi.22393

Strandén, I., and Garrick, D. J. (2009). Technical Note: Derivation of Equivalent Computing Algorithms for Genomic Predictions and Reliabilities of Animal merit. *J. Dairy Sci.* 92, 2971–2975. doi:10.3168/jds.2008-1929

Takahashi, A., Flanigan, M. E., McEwen, B. S., and Russo, S. J. (2018). Aggression, Social Stress, and the Immune System in Humans and Animal Models. *Front. Behav. Neurosci.* 12. doi:10.3389/fnbeh.2018.00056

Terenina, E., Babigumira, B. M., Le Mignon, G., Bazovkina, D., Rousseau, S., Salin, F., et al. (2013). Association Study of Molecular Polymorphisms in Candidate Genes Related to Stress Responses with Production and Meat Quality Traits in Pigs. *Domest. Anim. Endocrinol.* 44, 81–97. doi:10.1016/j.domaniend.2012.09.004

Tsuruta, S., and Misztal, I. (2006). "THRGIBBSF90 for Estimation of Variance Components with Threshold and Linear Models," in Abstract retrieved from 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil, August 13–18, 2006.

Tulloh, N. M. (1961). Behaviour of Cattle in Yards. II. A Study of Temperament. *Anim. Behav.* 9, 25–30. doi:10.1016/0003-3472(61)90046-X

USDA–APHIS, United States Department of Agriculture-Animal and Plant Health Inspection Service (2017). Beef Cow-Calf Management Practices in the United States. Available at: www.ers.usda.gov/publications/pub-details/?pubid=44532 (Accessed September, 2021).

Valente, T. S., Sant'Anna, A. C., Baldi, F., Albuquerque, L. G., and da Costa, M. J. R. P. (2015). Genetic Association between Temperament and Sexual Precocity Indicator Traits in Nellore Cattle. *J. Appl. Genet.* 56, 349–354. doi:10.1007/s13353-014-0259-0

Vallée Marcotte, B., Cormier, H., Rudkowska, I., Lemieux, S., Couture, P., and Vohl, M.-C. (2017). Polymorphisms in FFAR4 (GPR120) Gene Modulate Insulin Levels and Sensitivity after Fish Oil Supplementation. *Jpm* 7, 15. doi:10.3390/jpm7040015

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., et al. (2014). Genome-wide Association Mapping Including Phenotypes from Relatives without Genotypes in a Single-step (ssGWAS) for 6-week Body Weight in Broiler Chickens. *Front. Genet.* 5, 1–10. doi:10.3389/fgene.2014.00134

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide Association Mapping Including Phenotypes from Relatives without Genotypes. *Genet. Res.* 94, 73–83. doi:10.1017/S0016672312000274

Wang, M., Lai, X., Yu, H., Wang, J., Chen, Z.-Q., Lan, X.-Y., et al. (2010). Two Novel Single Nucleotide Polymorphisms (SNPs) and 4-bp Deletion Mutation of RBP4 Gene in Chinese Cattle. *J. Genet.* 89, 233–236. doi:10.1007/s12041-010-0031-0

Weikard, R., Hadlich, F., and Kuehn, C. (2013). Identification of Novel Transcripts and Noncoding RNAs in Bovine Skin by Deep Next Generation Sequencing. *BMC Genomics* 14, 789. doi:10.1186/1471-2164-14-789

Willham, R. L. (2000). Problems in Estimating Maternal Effects. *Livestock Prod. Sci.* 7, 405–418. doi:10.1016/0301-6226(80)90080-9

Zhang, X., Xu, Y. N., Chen, B., and Kang, L. (2020). Long Noncoding RNA PAHAL Modulates Locust Behavioural Plasticity through the Feedback Regulation of Dopamine Biosynthesis. *Plos Genet.* 16, e1008771–29. doi:10.1371/journal.pgen.1008771

Conflict of Interest: Author SM was employed by the Angus Genetics Inc.

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.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 July 2022
doi: 10.3389/fgene.2022.872486

Imputation of Ancient Whole Genome *Sus scrofa* DNA Introduces Biases Toward Main Population Components in the Reference Panel

J. A. M. Erven[1]*, *C. Çakirlar*[1], *D. G. Bradley*[2], *D. C. M. Raemaekers*[1] and *O. Madsen*[3]

[1]Groningen Institute of Archaeology, University of Groningen, Groningen, Netherlands, [2]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, [3]Animal Breeding and Genomics, Wageningen University and Research, Wageningen, Netherlands

Sequencing ancient DNA to high coverage is often limited by sample quality and cost. Imputing missing genotypes can potentially increase information content and quality of ancient data, but requires different computational approaches than modern DNA imputation. Ancient imputation beyond humans has not been investigated. In this study we report results of a systematic evaluation of imputation of three whole genome ancient *Sus scrofa* samples from the Early and Late Neolithic (~7,100–4,500 BP), to test the utility of imputation. We show how issues like genetic architecture and, reference panel divergence, composition and size affect imputation accuracy. We evaluate a variety of imputation methods, including Beagle5, GLIMPSE, and Impute5 with varying filters, pipelines, and variant calling methods. We achieved genotype concordance in most cases reaching above 90%; with the highest being 98% with ~2,000,000 variants recovered using GLIMPSE. Despite this high concordance the sources of diversity present in the genotypes called in the original high coverage genomes were not equally imputed leading to biases in downstream analyses; a trend toward genotypes most common in the reference panel is observed. This demonstrates that the current reference panel does not possess the full diversity needed for accurate imputation of ancient *Sus*, due to missing variations from Near Eastern and Mesolithic wild boar. Imputation of ancient *Sus scrofa* holds potential but should be approached with caution due to these biases, and suggests that there is no universal approach for imputation of non-human ancient species.

Keywords: imputation, ancient DNA (aDNA), *Sus scrofa*, animal husbandry, Neolithic

OPEN ACCESS

Edited by:
Xiangdong Ding,
China Agricultural University, China

Reviewed by:
Luca Ermini,
Luxembourg Institute of Health, Luxembourg
Huashui Ai,
Jiangxi Agricultural University, China

***Correspondence:**
J. A. M. Erven
jolijn_erven@hotmail.com

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

Received: *09 February 2022*
Accepted: *20 May 2022*
Published: *12 July 2022*

Citation:
Erven JM, Çakirlar C, Bradley DG, Raemaekers DCM and Madsen O (2022) Imputation of Ancient Whole Genome Sus scrofa DNA Introduces Biases Toward Main Population Components in the Reference Panel. Front. Genet. 13:872486. doi: 10.3389/fgene.2022.872486

1 INTRODUCTION

Recent advances in sequencing techniques led to a dramatic increase in the amount of retrievable ancient DNA (aDNA) from archaeological remains (Kircher, 2012), providing new insights into recent evolutionary history (Slatkin and Recimo, 2016; MacHugh et al., 2017; Brunson and Reich, 2019; McHugo et al., 2019). Poor preservation and contamination of exogenous DNA restricts sequence quality, reliability, and coverage of aDNA from archaeological bones (Pääbo et al., 2004; Prüfer et al., 2010). Furthermore, the damaged nature of aDNA poses computational challenges and introduces biases to the analysis of aDNA (Höss et al., 1996; Brotherton et al., 2007; Briggs et al.,

2009; Prüfer et al., 2010; Ginolhac et al., 2011; Sánchez-Quinto et al., 2012; Parks and Lambert, 2015; Kistler et al., 2017). One way to counter these problems is imputation, which is a powerful way to improve the quality of data and can potentially maximize the power of analysis that require dense genotypes such as runs of homozygosity (ROH), in depth admixture and trait association analyses (Gamba et al., 2014; Martiniano et al., 2017). Imputation is widely employed in studies of modern data (Van den Berg et al., 2019; Ye, et al., 2019), targeting allele frequencies from a set of reference individuals to infer allele frequencies at unknown or missing sites (Browning and Browning, 2007; Ausmees, 2019).

In aDNA studies, imputation has been applied on human genomes and achieved high levels of concordance between imputed genotypes and their high-quality (HQ) counterparts (>99%) (Gamba et al., 2014; Martiniano et al., 2017; Ausmees et al., 2021). Imputation of aDNA beyond humans is lacking; livestock aDNA is critical to understand pivotal moments in recent evolution such as domestication and pose an excellent case study. A number of factors are known to influence imputation ranging from reference panel characteristics to demographic history; assessing the potential and limitations of imputation of species beyond model species like humans is valuable to aid our understanding of not only imputation performance but also recent evolutionary events.

This paper assesses the power of imputation to increase the quality and information potential of low coverage aDNA samples, using *Sus scrofa* as a case study. This species is an intensively studied livestock species in terms of aDNA, particularly in the context of expansion of animal husbandry into Europe and significantly enhancing our understanding of how farming started in Europe (Larson et al., 2007; Ottoni et al., 2013; Frantz et al., 2019). Investigations have indicated that ancient Near Eastern domestic pigs lost their Near Eastern genomic signatures after their introduction to Europe (Larson et al., 2007; Frantz et al., 2019). Obtaining HQ samples to pinpoint the pace and nature of this turnover in different regions and shorter timescales in relation to larger societal and economic developments is necessary, but it remains a challenge due to poor preservation and contamination. To address this challenge, a systematic evaluation of different imputation methods was performed on whole genome ancient *Sus scrofa* DNA using data from a recent study consisting of ancient whole genomes of pigs sequenced to an appropriate depth for imputation (Frantz et al., 2019). Imputation achieved high genotype concordance but this is paired with biases toward a fraction of the reference panel. These biases might be related to the size and diversity of the reference panel, the reference genome, or the genetic architecture of pigs, and they impose limitations on the interpretive power of imputed data in terms of the proposed genomic turnover of this species in particular and in general the evolution of animal husbandry in Neolithic Europe.

2 MATERIALS AND METHODS

Evaluating imputation of *Sus scrofa* aDNA by comparing three tools, two pipelines, and three variant calling methods.

2.1 Data Description and Preparation

2.1.1 Ancient Samples

Seven archaeological samples with high-coverage data and four archaeological samples with moderate coverage from Frantz et al. (2019) were used (**Table 1**; **Supplementary Table S1**). Raw FASTQ reads were downloaded from the ENA (accession numbers see **Supplementary Table S1**). Raw reads were trimmed using *cutadapt v2.10* (Martin, 2011) for quality (<20), length (<20) and adapters used in the library preparation (Meyer and Kircher, 2010). *FastQC v0.11.9* quality reports were made for the raw and trimmed data (Andrews, 2010). The trimmed reads were aligned applying the *Burrows-Wheeler algorithm (BWA) aln v0.7.17* (Li and Durbin, 2009) to the *Sus scrofa* 11.1 reference genome (Warr et al., 2020), with default parameters apart from disabling the seed option (–l 1024), increasing the maximum number of gap opens (–o 2) and changing the maximum edit distance (–n 0.01). Duplicates were removed with *Picard MarkDuplicates v2.18.17* (http://broadinstitute.github.io/picard) and BAM files from different sequencing lanes were merged using *SAMtools merge v0.1.19* (Li et al., 2009). Duplicates were removed with *FilterUniqueSamCons*.py for the merged BAM files (Kircher, 2012). Indels were realigned with *GATK 3.8 RealignerTargetCreator* and *IndelRealigner* with default parameters (Van der Auwera et al., 2013). Depth of coverage and quality were computed using *Qualimap v2.2.1* (Okonechnikov et al., 2015). Molecular damage was assessed using *MapDamage2.0* using default parameters (Jónsson et al., 2013).

Contamination from prokaryotes and humans was assessed by calculating percent identity score and coverage per read with *BLAST + Blastn Megablast v2.10.1* on prokaryotes, human and *Sus scrofa* databases (Camacho et al., 2008). Reads were considered contaminants if the percent identity (E-value) and coverage of the contaminants (prokaryotes and humans) was higher than the percent identity and coverage of *Sus scrofa*. Contaminated reads were removed from the BAM file with a custom-made python script.

Imputation was assessed by comparing imputed genotypes to their corresponding HQ genotypes, similar to previous studies (Gamba et al., 2014; Martiniano et al., 2017; Ausmees et al., 2021). Three of the seven samples with high-coverage data (KD033, KD037, and VEM185) were downsampled with *Picard v2.18.17* (http://broadinstitute.github.io/picard), to create low coverage samples for imputation ranging from 0.5 to 2× with steps of 0.5×. Three methods were used to assess the accuracy of imputation: Method 1, imputation with variant sites; Method 2, imputation with all confident sites; and Method 3, added to achieve higher genotype concordance which called only genotypes present in the reference panel. HQ genotypes were created from the high-coverage samples to create a golden standard. Genotype likelihoods were called with the *Genome Analysis Toolkit (GATK) UnifiedGenotyper v3.8.0* (Van der Auwera et al., 2013) using either each alignment data of the ancient samples individually or by joined SNP calling. Genotype likelihoods were called with a minimum quality of 25, with output mode *EMIT_VARIANTS_ONLY* for Method 1, *EMIT_ALL_*

TABLE 1 | Sample information. ID, origin, period and ancestry taken from Frantz et al. (2019).

ID	Origin	Period	Ancestry	Genome coverage
KD033	Germany-Herxheim	Neolithic	~46% European, ~54% Near Eastern	6.9
KD037	Germany-Herxheim	Neolithic	~91% European, ~9 Near Eastern	21.6
VEM185	England-Durrington Walls	Neolithic/Bronze Transition	~90% European, ~10 Near Eastern	21.7

TABLE 2 | Reference panels with their respective number of individuals/population.

Reference panel	Number of individuals
Main references	
Dutch wild boar-European wild boar (EUW)	12
Italian wild boar-European wild boar (EUW)	6
French wild boar-European wild boar (EUW)	1
Pig breeds-European Domestic (EUD)	25
Greek wild boar (BLW)	4
Near Eastern + Turkish wild boar- Near Eastern wild boar (NEW)	3
Total	51
Main + ancient references	
Main reference	51
Near Eastern-Ancients (ANC)	5
European-Ancients (ANC)	3
Total	59

CONFIDENT_SITES for Method 2 and output mode EMIT_ALL_SITES and genotyping mode GENOTYPE_GIVEN_ALLELES for Method 3, with -alleles genotypes from the reference panel. Variants were filtered to keep only autosomal, biallelic SNPs, and a minimum quality of 30. In order to avoid introducing a possible bias from nucleotide misincorporations due to post-mortem damage, the generated VCF (Variant Call Format) files were filtered to exclude all sites where the most likely genotype could have been inferred from a deaminated allele with a custom-made python script. For C→T deaminations, C↔T SNPs were excluded from further analyses if the most likely genotype contained a T allele, and for G→A deaminations, G↔A SNPs were excluded from further analyses if the most likely genotype contained an A allele. Genotypes were not filtered in Method 3 when using GLIMPSE, because this software only imputes genotypes present in the target VCF, they were instead kept as no calls (./.).

2.1.2 Reference Panel

The reference material used for imputation consisted of the wild boar and pig breeds collection of Wageningen University and two Iberian samples from Ramírez et al. (2015) (**Supplementary Table S1**). Pig breeds that have no known introgression with Asian breeds were selected to avoid potential bias. Genotype likelihoods were called with the *Genome Analysis Toolkit (GATK) UnifiedGenotyper v3.8.0* (Van der Auwera et al., 2013), with a minimum quality of 15, calling SNPs, with the mode *EMIT_VARIANTS_ONLY* for Method 1 and *EMIT_ALL_CONFIDENT_SITES* for Method 2. The reference panel was filtered to only include autosomal biallelic SNPs, a minimum quality of 30, and a minimum depth of 4, a call rate of 0.8, and removal of repetitive elements. In order to evaluate the effect of the reference composition on the imputation, multiple reference panels were considered. The main reference panel consists of modern pig breeds, European wild boar, and Near Eastern wild boar (51 individuals, 12,737,362 variants—**Table 2**; **Supplementary Table S2**). To deduce the effect of ancient samples on imputation, eight ancient individuals were added to the main reference panel, consisting of two Near Eastern samples, three ancient Near Eastern, and three ancient European samples (59 individuals, 10,823,257 variants—**Table 2**, **Supplementary Table S2**), called Main + ancient reference. Moreover, the main reference panel was divided into several subsets to pinpoint the effect of reference bias on imputation (See **Supplementary Material**-Subsets of reference panel). Additionally, to deduce the effect of Asian haplotypes on imputation, Asian wild boars, Asian domestic pigs and South-East Asian *Sus* were added to the reference panel (See **Supplementary Material**-Including Asian samples). Different filters and combinations of filters were used on the reference panel to optimize the imputation workflow and deduce the effects of these filters on imputation. These filters consisted of removing 1) transversions, 2) transitions, 3) filtering for minor allele frequency (MAF) bins {<0.05, 0.05–0.1, 0.1–0.3, >0.05, >0.3, No MAF}, and their various combinations. Results of all combinations can be found in **Supplementary Table S2**. The reference panels were phased with *Beagle5* (Browning et al., 2018), using default parameters apart from changing the effective population size (Ne) to 20,000 (Groenen et al., 2012).

2.2 Genetic Map

A genetic map was created using the recombination frequencies that Johnsson et al. (2020) estimated based on nine genotyped pedigrees on the *Sus scrofa* 11.1 reference genome. These

recombination frequencies were converted to cM using the Haldane formula (Haldane's Mapping Function, 2008). Genetic maps were made for each chromosome in the plink format with bins of 1 MB (**Supplementary Table S3**).

2.3 Imputation

For Methods 1–3 imputation was performed using *Impute5* and *Beagle5*, using default parameters, with a phased reference panel (**Supplementary Table S2**), with a *Ne* of 20,000 and, --div-select and –out-gp-field parameters for *Impute5* (Rubinacci et al., 2020) and window = 40, overlap = 4 and gp = true parameters for *Beagle5* (Browning et al., 2018). Imputation was performed for chromosome 1–18, individually and using sliding windows (See **Supplementary Material**—Chromosomal imputation). The effect of including multiple ancient samples on imputation was evaluated by imputing joint ancient samples and was compared to individual imputation (**Supplementary Material**—Joined Imputation). The focus was on individual imputation. Imputation was performed using two different imputation pipelines: 1) the original one-step pipeline used in Ausmees et al. (2021) and 2) the two-step pipeline used for low coverage samples in Hui et al. (2020). The two-step pipeline adds another filtering step prior to imputation that accounts for genotype probability. *Beagle 4.1* was used to calculate genotype probabilities for the target downsampled VCF using default parameters, with the same phased reference panel that was used for imputation (**Supplementary Table S2**), with a Ne of 20,000 and gprobs = true parameters. Variants with a genotype probability (GP) < 0.99 were removed from the target downsampled VCF, leaving only confident genotype calls. The imputed genotypes were filtered for an imputation score of 1 (highest imputation accuracy). For Method 3, *GLIMPSE v1.1.1* (Rubinacci, et al., 2021) was also used with similar settings as applied in ancient human imputation and the pipeline proposed by Rubinacci, et al. (2021), with default parameters, and a phased reference panel (**Supplementary Table S2**). Variants with an imputation score of <1 were removed from the target downsampled VCF, leaving only confident imputed genotype calls. GLIMPSE *v1.1.1* was only tested with Method 3 because of the incompatibility with the other two methods/pipelines.

2.4 Genotype Concordance

Imputation accuracy was assessed by genotype concordance defined as the fraction of genotypes that were imputed correctly. This was measured by dividing the incorrectly imputed SNPs with all imputed SNPs and was measured separately for each sample. The correctly and incorrectly imputed SNPs were derived from comparing the imputed SNPs to their HQ counterpart similar to the approach of *Picard GenotypeConcordance*. The HQ genotypes used are pre-deamination filtered, to keep confident transitions (transitions that also occurred in the reference panel), and not transitions arising from deamination. The incorrectly imputed SNPs were classified into incorrect positions (positions not occurring in HQ) and incorrect genotypes (genotypes different from HQ genotypes). Information content, that is, the amount of gained genotypes, was calculated by dividing the amount of imputed genotypes to the total amount of HQ genotypes.

2.5 Downstream Analysis

Downstream analyses were performed to investigate the difference and/or similarity between imputed and HQ genotypes. Data were pruned with *PLINK 1.9* (Purcell et al., 2007) with the parameters—geno 0.10. A principal component analysis (PCA) was performed on diploid genotypes consisting of the reference panel, the HQ samples and the imputed samples using *PLINK 1.9* pca on autosomes only. Eigenvalues and vectors were plotted with the use of *Mathplotlib* (Hunter, 2007) and *Seaborn* (https://zenodo.org/record/883859#.XSdFFugza01). An admixture analysis was performed using the same dataset as the PCA analysis, however separately for downsampled, imputed and HQ genotypes. A*DMIXTURE v1.3.0* (David et al., 2009) was used with standard parameters and K ranging from 2 till 5. Furthermore, bootstrapping was performed using the parameter -B. Identical By Descent (IBD) analysis was conducted on the same dataset as the admixture analysis. *IBDseq v2.0* with standard parameters was used to calculate IBD segments between samples (Browning and Browning, 2013). A regions of homozygosity analysis was performed using the same dataset as the admixture analysis using plink –homozyg with the parameters –homozyg-kb 10, --homozyg-gap 10, --homozyg-snp 100, --homozyg-window-het 2, --homozyg-window-snp 100 --homozyg-window-missing 1. *DetectRUNS* (https://cran.r-project.org/package=detectRUNS) was used to visualize and calculate ROH statistics.

2.6 Reference Affinity

Correct and incorrect imputed genotypes were compared to their HQ counterpart to assess whether imputed genotypes show a systematic bias toward the reference genome. Reference bias was measured as the presence/absence of different ancestral/origin groups between the correct and incorrect imputed genotypes and their HQ counterpart.

3 RESULTS

Genotype concordance was calculated for three imputation tools, two pipelines and three variant calling methods to test the best method to approach imputation in *Sus scrofa*. Downstream analyses were performed to assess the accuracy and power of imputation.

3.1 Genotype Concordance

3.1.1 Tools: Beagle5 Versus Impute5

Genotype concordance was higher for Beagle5 compared to Impute5 for KD037 and VEM185 but lower for KD033 (**Figure 1**). For both tools, KD037 and VEM185 performed better than KD033, this being more pronounced for Beagle5. The amount of correctly imputed variants differed greatly between the tools, with Beagle5 being systematically lower (**Figures 1IC,D**). Impute5 imputed 25%–34% of the total amount of HQ genotypes, while Beagle5 imputed around 5%.

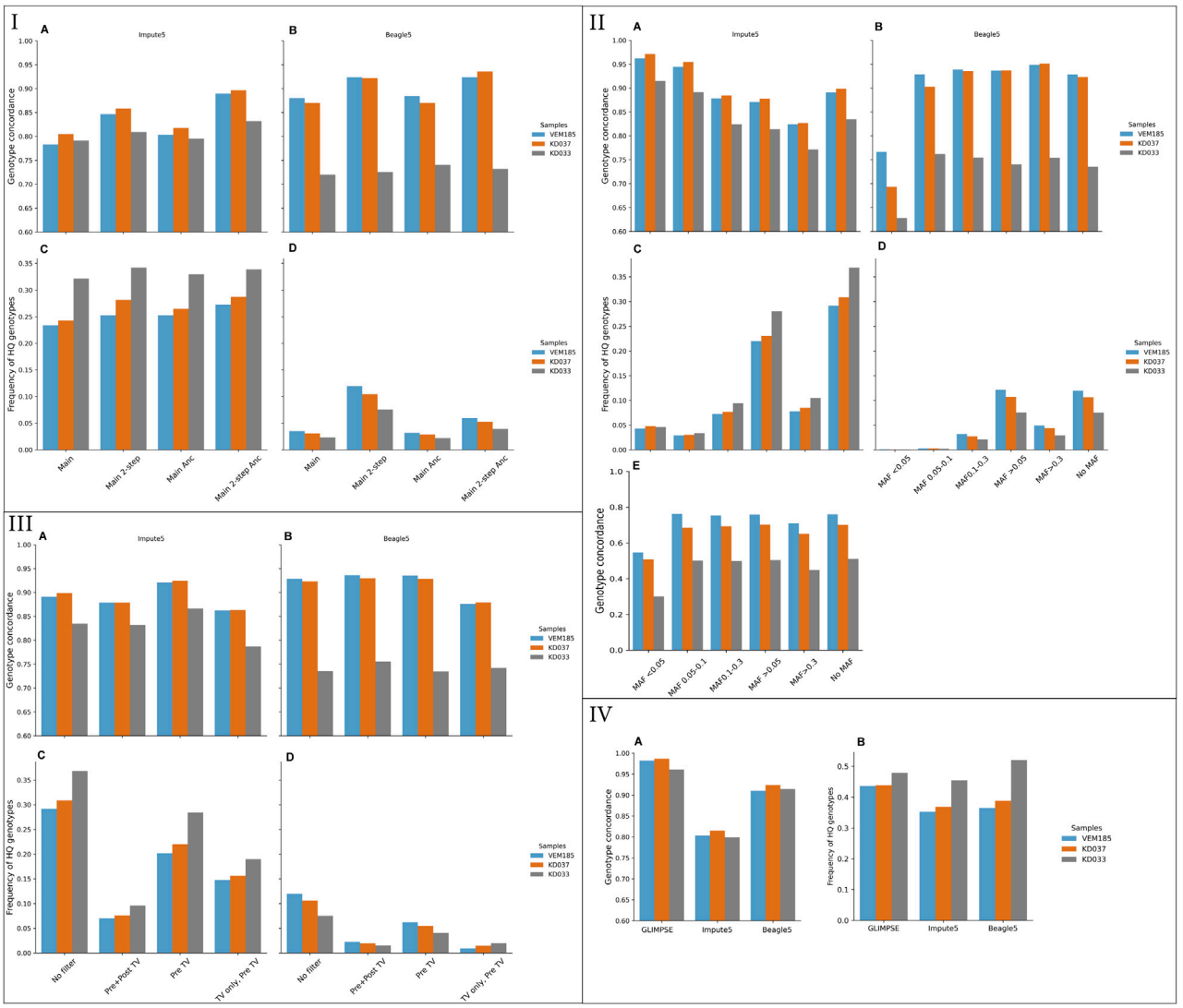


FIGURE 1 | Genotype concordance and fraction of high-quality (HQ) genotypes (portrayed in decimals, total HQ genotypes: VEM185: 4,701,683, KD037: 4,531,126, KD033: 3,887,848) using different imputation tools and reference panels. Imputed from 1× downsampled coverage genomes of VEM185, KD037, and KD033. I) Genotype concordance for Impute 5 **(A)**, Beagle5 **(B)**, Fractions of HQ genotypes covered by imputation in Impute5 **(C)**, and Beagle5 **(D)**. II) Genotype concordance and Fraction of HQ genotypes of various MAF bins of Main + 2-step + Anc + all reference panel. Genotype concordance for Impute 5 **(A)**, Beagle5, **(B)** Fractions of HQ genotypes covered by imputation in Impute5 **(C)**, Beagle5 **(D)** and, Genotype concordance of only heterozygotic variants of various MAF bins imputed with Impute5 **(E)**. III) Filters on target downsampled vcf. Genotype concordance for Impute 5 **(A)**, Beagle5 **(B)**, Fractions of HQ genotypes covered by imputation in Impute5, **(C)** Beagle5 **(D)**. IV) Genotype concordance **(A)** and fraction of HQ genotypes **(B)** from GLIMPSE, Impute5 and Beagle5 for Method 3. * Genotype concordance of Method 3 was done on all sites and cannot be directly compared to the other methods. Main, original one-step pipeline on main reference panel; 2-step, two-step pipeline with an extra filtering step; Anc, Ancient samples included in the reference panel; TV, transversions.

Beagle5 achieved the highest genotype concordances in KD037 and VEM185, but produced less correctly imputed variants. Impute5, on the other hand, achieved the highest genotype concordance in KD033 and produced more correctly imputed variants. Furthermore, genotype concordance between chromosomes was more uniform in Beagle5 compared to Impute5 (**Supplementary Table S4**). These results are based on the default one-step pipeline. The main one-step pipeline was extended to the two-step pipeline to test various settings for both tools that could influence imputation accuracy changing one element at a time.

3.1.2 Pipeline: One-Step Versus Two-Step

Genotype concordance in the two-step pipeline was higher compared to the one-step pipeline for both tools (**Figures 1IA,B**). The amount of correctly imputed variants increased for Beagle5 whereas the amount of correctly imputed variants for Impute5 stagnated (**Figures 1IC,D**). Genotype concordance between chromosomes was more uniform in

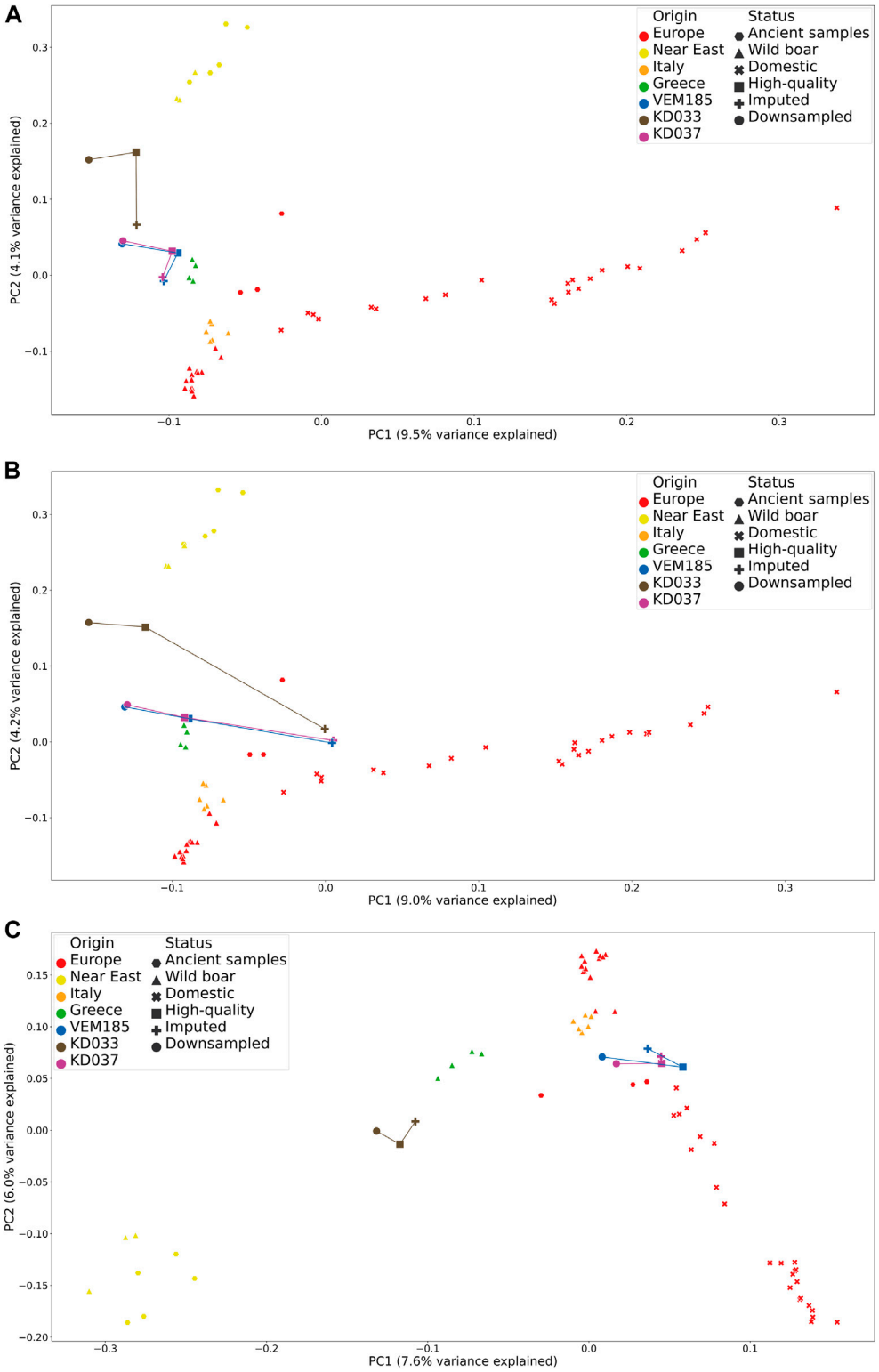


FIGURE 2 | Principal component analysis comparing high-quality, imputed genotypes and downsampled data together with samples from the reference panel. IMP1 **(A)**, IMP4 **(B)**, and IMP5 **(C)**.

the two-step pipeline compared to the one-step pipeline (**Supplementary Table S4**) (variation of filter combinations used for the comparisons can be found in **Supplementary Table S2**).

3.1.3 Reference Panel: With and Without Ancient Samples

Genotype concordance was ~5% higher in the two-step pipeline when using the reference panel including ancient samples for Impute5, whereas the inclusion of ancient samples only provided a 0.5% different genotype concordance for Beagle5 (**Figures 1IA,B**; **Supplementary Table S4**). Similarly, genotype concordance between chromosomes showed more uniformity with ancient samples than without ancient samples for Impute5, but not for Beagle5 (**Supplementary Table S4**). The amount of correctly imputed variants with respect to the inclusion of ancient samples had no effect for Impute5 but decreased for Beagle5 (**Figures 1IC,D**).

3.1.4 Reference Panel: Variant Sites Versus All Confident Sites Category (All)

Using the *all confident sites* category, method 2, slightly increased genotype concordance for both tools (**Supplementary Figures S1A,B**). The amount of correctly imputed variants was larger in the *all confident sites* category compared to the *variant sites* category, method 1, for both tools (**Supplementary Figures S1C,D**). The uniformity of genotype concordance between chromosomes was more equal in the all confident sites category compared to the variant sites category for Beagle5, except for KD037 (**Supplementary Table S4**).

3.1.5 Reference Panel: Pre-Imputation Filters Versus Standard

Reference panels filtered for transversions and transitions showed similar genotype concordance, with transversions only having the lowest genotype concordance (**Supplementary Figures 2A,B**). The amount of correctly imputed variants decreased drastically for only transitions and only transversions with roughly 50%, in both tools (**Supplementary Figures 2C,D**). The uniformity of concordance between chromosomes was equal for all filters (**Supplementary Table S4**).

Reference panels filtered for MAF showed variation in genotype concordance, where MAF bins <0.05 and 0.05–0.10 had the highest and MAF >0.3 the lowest genotype concordance for Impute5 (**Figure 1IIA**). This contrasted with Beagle5, where MAF <0.05 had the lowest genotype concordance and MAF >0.3 the highest (**Figure 2B**). Filtering for MAF (Beside the common variant >0.05 filter) drastically decreased the amount of correctly imputed variants in both tools (**Figures 1IIC,D**). Genotype concordance of heterozygotes did not show the same trend as all variant genotype concordance for Impute5 (**Figure 1IIE**), ~30% of the total imputed genotypes were heterozygotes (**Supplementary Table S4**). Genotype concordance of MAF bin <0.05 decreased drastically, while the other MAF bins decreased more modestly.

3.1.6 Target VCF: Filters Versus No Filters

The target VCF was filtered pre- and post-imputation to deduce the effect on genotype concordance. The reference panels used for this analysis were IMP1 and IMP2. Filtering for transitions pre-imputation, thus only keeping transversions, had the most favorable effect on genotype concordance. The highest genotype concordance for Beagle5 was filtering for transitions pre- and post-imputation, whereas the highest genotype concordance for Impute5 was filtering for transitions pre-imputation (**Figures 1IIIA,B**). The amount of correctly imputed variants of the pre- and post-imputation filters decreased drastically compared to using no filter on the target VCF for both tools (**Figures 1IIIC,D**). The imputed variants were further filtered for well-known positions, namely the 50 k porcine SNP-Chip (Yang et al., 2017), the transversions only, and main SNP-sets from the study of Frantz et al. (2019). These filters did not improve genotype concordance and decreased the amount of correctly imputed variants (**Supplementary Figure S3** — Methods Known genomic positions).

TABLE 3 | Reference panel abbreviations.

ID	Reference panel
IMP1	Main + 2-step + Ancients + All confident sites, Impute5
IMP2	Main + 2-step + Ancients + All confident sites, Beagle5
IMP3	Main + 2-step + Ancients + All confident sites + MAF<0.05, Impute5
IMP4	Main + 2-step + Ancients + All confident sites + MAF>0.3, Beagle5
IMP5	Main + 2-step + Ancients + All, GLIMPSE

3.1.7 Target VCF: All Sites (Sites Present in Reference Panel)

A method that has been shown to achieve high genotype concordance in ancient human imputation (Martiniano et al., 2017; Hui, et al., 2020; Method 3) was also applied. This method was tested on three imputation tools, Beagle5, Impute5, and GLIMPSE. GLIMPSE achieved the highest genotype concordance, reaching 98% in KD037 and VEM185 and 96% in KD033 (**Figure 1IVA**). Genotype concordance for Beagle5 stayed constant for VEM185 and KD037 but increased for KD033 compared to the two-step pipeline. Genotype concordance for Impute5 decreased for all samples compared to the two-step pipeline. The amount of correctly imputed variants increased for all samples and all tools, reaching roughly 50% (**Figure 1IVB**). Furthermore, the genotype concordance between chromosomes was more constant for this method compared to the two-step pipeline (**Supplementary Table S4**).

3.2 Downstream Analysis

Downstream analyses were performed on the imputed genotypes IMP1, 2, 3, 4, and 5 (Full descriptions found in **Table 3**). PCA were performed to pinpoint and compare the genetic affinities of imputed, HQ and downsampled samples. Variation captured by the first two principal components of IMP1 shows that the imputed genotypes of KD033, KD037, and VEM185 cluster closer to their HQ counterparts in the first principal component (PC1, 9.5% variation) but tend to have a bias toward European wild boar in the second principal component (PC2, 4.1% variation) (**Figure 2A**). This bias is greater for KD033 compared to KD037 and VEM185. PCA of IMP2 shows that the downsampled genotypes and the HQ

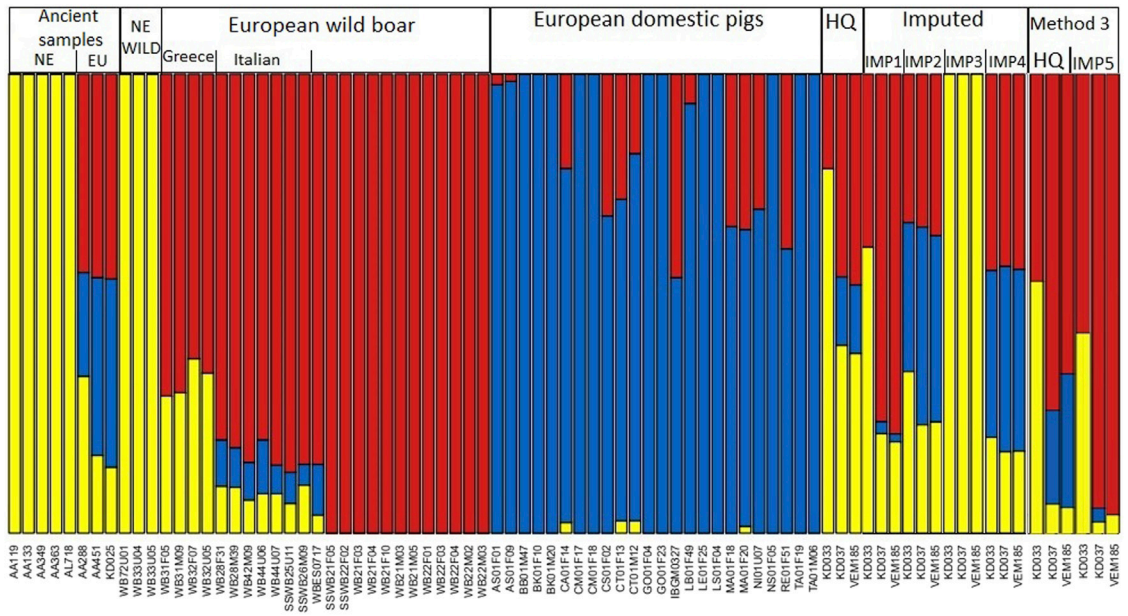


FIGURE 3 | Admixture analysis with K = 3, comparing high-quality and imputed data from IMP1-5 together with samples from the reference panel. NE, Near East; EU, European. The colors represent the three ancestral groups, where blue is European domestic pig, red is European wild boar, and yellow is Near Eastern.

genotypes cluster closer to each other than to the imputed genotypes, the latter showing a bias toward the domestic pigs in the first principal component (PC1, 9% variation) (**Supplementary Figure S4**). PCA of IMP3 shows that rare alleles have a bias toward the domestic cluster in the first principal component (PC1, 9% variation) and a domestic and European wild boar bias in the second principal component (PC2, 3.8% variation) (**Supplementary Figure S5**). The PCA for IMP4 has the same trend as IMP2, where the downsampled genotypes and the HQ genotypes cluster closer to each other than to the imputed genotypes. However, the imputed genotypes of KD037 and VEM185 show a decreased bias toward the domestic pigs on the second principal component (PC2, 4.2% variation) (**Figure 2B**). The PCA for IMP5 shows a bias toward the European wild boar cluster for KD033 and VEM185 in both principal components, where the imputed genotypes of VEM185 cluster closer with the HQ genotypes than the downsampled genotypes (**Figure 2C**). This trend is not observed in KD033. The imputed genotypes of KD037 cluster closely toward the HQ counterpart, showing a slight bias on PC2 toward the European wild boar cluster. However, there seems to be a slight bias introduced in the HQ and downsampled genotypes, which is more evident for KD037 and VEM185. They are clustering more toward the European domestic cluster than in their previous PCA (IMP1–4). Beagle5 showed a similar trend as GLIMPSE but with an elevated bias toward the downsampled genotypes (**Supplementary Figure 6SA**). Impute5 showed an increased amount of bias toward the European wild boar cluster (**Supplementary Figure 6SB**).

Admixture analysis of three ancestral groups (K = 3) shows the genetic ancestry of reference panel and HQ samples and indicates a presence of all three ancestral groups in KD037 and VEM185, and two ancestral groups in KD033. The ancestral groups of KD037 and VEM185 are similar, with VEM185 having a slightly larger "European domestic pig" component, whereas KD033 consists of a larger Near Eastern and smaller European component. Admixture analysis of IMP1 shows an increase in the European component, and a decrease of the other components of all imputed samples, highlighting a bias toward European samples. Admixture analysis of IMP2 shows a decrease of the European component in KD037 and VEM185 and an increase in KD033. The most noticeable difference between IMP2 and HQ is the increased component of "European domestic pigs" in all three imputed samples, again highlighting a potential bias toward "European domestic pigs." KD033 showed most bias losing almost half of its Near Eastern component and increasement of both European wild boar and domestic pig's components. Admixture analysis of IMP3 shows only one ancestral component, namely, the Near Eastern component (**Figure 3**). This potential bias toward the Near Eastern component could have arisen from a low amount of variants present in the imputed genotypes. Admixture analysis of IMP4 is similar to IMP2, with a slight decrease in Near Eastern ancestral component and an increase in European ancestral component, showing a bias toward the "European domestic pigs" component in all samples and a bias toward the European wild boar component in KD033. Admixture analysis for Method 3 shows a deviation between the two HQ, where the HQ in Method 3 shows a decrease in Near Eastern components and an increase in European wild and domestic components. Therefore, IMP5 was compared to the HQ of Method 3. Admixture analysis of IMP5 shows an increase in the European wild boar component in all samples and a decrease of "European domestic pig" and Near Eastern component in

TABLE 4 | ROH statistics of the IMP1 reference panel for all autosomes shown per class of 0–0.5 mb, 0.5–1 mb, and >1 mb.

	0–0.5			0.5–1			>1		
	Count	**Sum kb**	**Froh**	**Count**	**Sum kb**	**Froh**	**Count**	**Sum kb**	**Froh**
KD033_HQ	4,496	232,157	0.1025	1	516	0.0002	NA	NA	NA
KD033_Imp1	3,287	343,670	0.1517	17	13,152	0.0058	2	2,377	0.0010
KD037_HQ	4,434	193,928	0.0856	NA	NA	NA	NA	NA	NA
KD037_Imp1	2,743	285,325	0.1259	12	8,626	0.0038	1	1404	0.0006
VEM185_HQ	3,071	124,173	0.0548	NA	NA	NA	NA	NA	NA
VEM185_Imp1	2,585	266,071	0.1174	14	9,909	0.0044	1	1600	0.0007

Each class has three statistics, ROH count, total sum of kb and Froh.

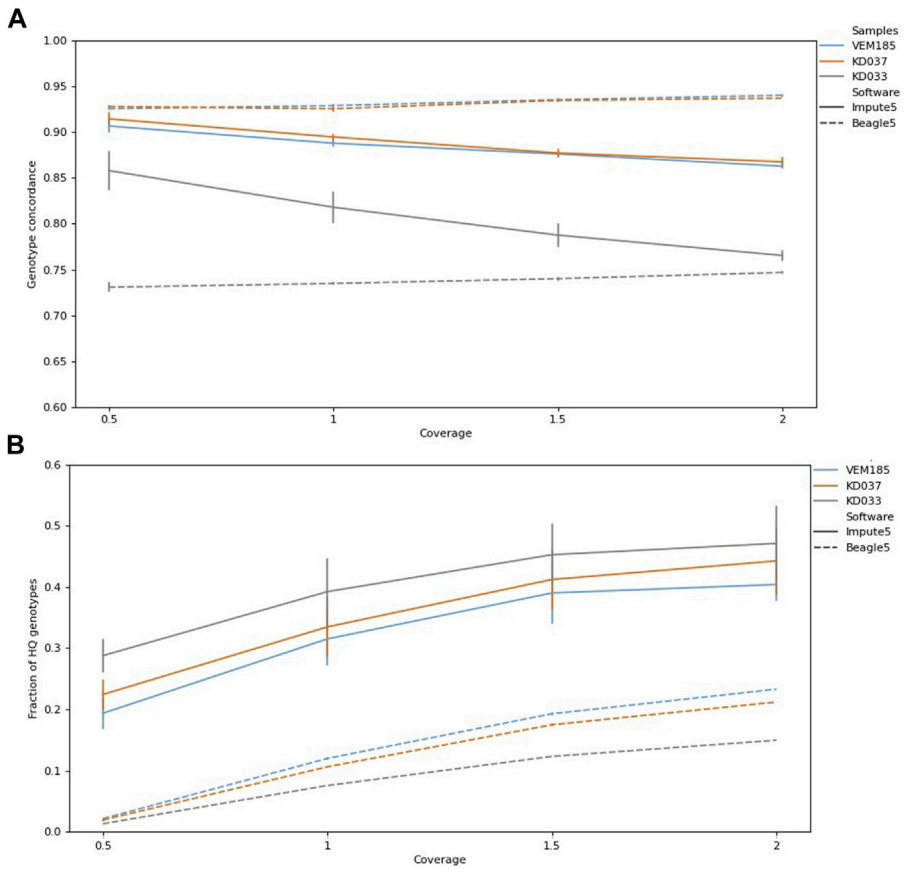


FIGURE 4 | Genotype concordance and fraction of high-quality (HQ) genotypes (total HQ genotypes: VEM185:4,701,683, KD037: 4,531,126, KD033, 3,887,848) of imputed samples for different levels of coverage for Impute5 and Beagle5. Imputation was performed using IMP1 and IMP2. Bars are standard errors. Genotype concordance **(A)**, Fraction of HQ genotypes **(B)**.

KD037 and VEM185, whereas KD033 only has a decrease in the Near Eastern component.

Identity-By-Descent (IBD) analysis shows that the imputed genotypes of IMP1 share large IBD segments with each other, covering whole chromosomes. These IBD segments are not present in their HQ or downsampled counterparts. IMP2 showed more variation with the imputed genotypes resulting in more fragmented IBD segments when compared to IMP1. However, most of the IBD segments do not overlap with the HQ IBD segments. IMP3 and 4 did not have enough depth to perform a proper IBD analysis. ROH analysis shows a similar but less drastic trend. The amount of ROHs was smaller in the imputed samples but they consisted of longer stretches (**Table 4**). The imputed samples had considerably larger ROHs, some larger than 1 MB, while the HQ samples had smaller fragmented ROHs. The elongated ROH stretches in the imputed samples attributed to a higher Froh compared to the HQ samples (**Table 4**). However, the ROHs in the imputed samples overlap with the HQ samples

but consists of longer stretches (**Supplementary Appendix** ROH). The ROH analysis was only performed for IMP1, because Beagle5 had a low amount of variants.

3.3 Effects of Coverage on Imputation

Coverage levels vary in genotype concordance, reaching 0.94 for KD037 and VEM185, and 0.75 for KD033 using Beagle5, where 2× reached the highest genotype concordance (**Figure 4A**). This trend is opposite for Impute5, which reached a genotype concordance of 0.92 for KD037, 0.91 for VEM185, and 0.88 for KD033, with the lowest coverage 0.5×. Imputed genotypes increased with increasing coverage (**Figure 4B**)

3.4 Effects of Reference Bias

The HQ genotypes were considered a baseline of the true genotypes that overlap with the groups in the reference panel (**Supplementary Figure S7**). For Impute5, the correctly imputed genotypes showed a bias toward the genotype that is most common in the reference panel and occurs across all groups (Ancients, EUW, EUD, and NEW), while the incorrectly imputed genotypes showed a bias toward the EUD; EUW and EUW groups (**Supplementary Figures S8,9**). For Beagle5, the correctly imputed genotypes showed a bias toward ANC; EUD; EUW, EUD; EUW, EUD; EUW; NEW, and, ANC; EUD, which is similar to the bias shown in the incorrectly imputed genotypes (**Supplementary Figures S10,11**). The incorrectly imputed genotypes were randomly divided throughout the chromosomes (**Supplementary Figure S12**).

4 DISCUSSION

The analyses revealed that for imputation of *Sus scrofa* aDNA data: 1) genotype concordance is relatively high, similar to modern imputation, with a minor increase in information content (fraction of gained genotypes in relation to HQ) in imputed genotypes and 2) imputation performance showed inaccuracies in downstream analyses. These results have a variety of implications for our understanding of the potency of imputation of non-human ancient DNA in terms of its performance and limitations.

4.1 Imputation Performance

The relatively high genotype concordance of 0.95 for Beagle5, 0.925 for Impute5, and 0.98 for GLIMPSE at 1× coverage on ancients is along the lines of genotype concordance in imputation of modern breeds (see Song et al., 2019; Ye et al., 2019; Wang et al., 2021). The higher genotype concordance in KD037 and VEM185 compared to KD033 might be explained by their difference in ancestry. Frantz et al. (2019) have shown that KD033 possessed ~54% Near Eastern ancestry, while KD037 and VEM185 possessed only ~10% Near Eastern ancestry (Frantz et al., 2019). The larger component of Near Eastern ancestry in KD033 may have caused the lower performance due to the reference panel being skewed toward European individuals. Another explanation could be the difference in coverage between the samples, KD037 and VEM185 both have coverages >20×, whereas KD033 has a coverage of ~7x. However, KD037 and VEM185 had the same genotype concordance when downsampled to a similar coverage, excluding this possibility (**Supplementary Methods**- Downsampling KD037 and VEM185). Moreover, KD033 showed most deviation when downsampled multiple times, showing that the ancestry components of this sample seem to be a factor in the level of accuracy in imputation.

All tools achieved high genotype concordance but differed in amount of information gained. Moreover, Beagle5 showed less variation in imputation of repeated downsampled VCFs compared to Impute5, showing that Beagle5 might be less affected by the randomness of downsampling. Genotype concordance increased with the two-step imputation pipeline. This was specifically designed for genomes with low coverage (Hui et al., 2020). Our results indicate that non-model species and species without an extensive reference panel could also benefit from this approach. Furthermore, genotype concordance increased when ancient samples were added to the reference panel, adding to the number and diversity of individuals. Finally, genotype concordance and number of correctly imputed genotypes increased when using all confident sites but this increased computational time and memory significantly. GLIMPSE achieved the highest genotype concordance with Method 3, that consisted of reference panel called genotypes in the three target downsampled samples. However, this method did not improve the genotype concordance for Impute5 and Beagle5, but did improve amount of genotypes recovered in all tools.

When only looking at genotype concordance the imputation performance of imputation of *Sus scrofa* aDNA could be deemed sufficient. However, there are potential shortcomings. High genotype concordance obtained in the imputed genotypes does not result in an equal representation of genotypes from the original high coverage genome and consists of only a subset, covering roughly 5%–50%. Moreover, imputed genotypes showed greater affinity with populations that are overrepresented in the reference panel as seen in downstream analyses (e.g., PCA, Admixture). One example of the unequal representation of genotypes from the original HQ genome is apparent from genotype concordances on different MAF bins. Genotype concordance in rare alleles (MAF < 0.05) reached 97% but resulted in a bias toward main components in the reference panel in downstream analyses. This is potentially due to the representation of only ~5% of the original HQ genotypes in the imputed genotypes It is therefore essential to look beyond genotype concordance and focus on multiple aspects like fraction of HQ genotypes obtained by imputation and potential biases in downstream analysis.

Downstream analyses can identify how imputed genotypes act in comparison to their HQ counterpart. The PCA of IMP1 resulted in accurate clustering of imputed and HQ genotypes with only a slight bias toward the European wild boar component. This same analysis for Beagle5 resulted in a stronger bias toward the European domestic pig component. This illustrated that a high genotype concordance does not necessary lead to accurate downstream analyses. The imputed genotypes are correct, but introduce bias in subsequent downstream analyses because they are from specific regions of the genome and are not informative enough to detect genetic variation among samples. This trend is also apparent in the admixture analysis, where imputed genotypes have biases toward European wild boar and domestic pig components. IMP3 is an exception that might be

attributed to the high amount of missing genotypes pulling it toward the Near Eastern component that featured missingness, due to low coverage and ancient individuals. IMP5 achieved the highest genotype concordance and resulted in the most accurate clustering for KD037 in the PCA, suggesting that imputation of ancient *Sus* is feasible. However, VEM185 had a similar genotype concordance as KD037 but showed the most bias in downstream analyses for this specific method, implying that high genotype concordance does not preclude bias across samples.

The IBD analysis shows that imputed genotypes of different samples from Impute5, share large IBD segments, sometimes even stretching chromosome wide. This could be a result of: 1) samples which lost their individual variation and became more similar due to imputation and/or 2) imputed genotypes that did not have enough depth for IBD analysis. The second explanation is unlikely, as imputed genotypes for Beagle5, did not show these large IBD segments. The ROH analysis shows that there are longer homozygous stretches throughout the genome in imputed genotypes compared to their HQ counterparts. Causes for this could be that the imputed genotypes were predominantly homozygous with little representation of heterozygotes, contributing to long ROHs and that the imputed genotypes have less markers than the HQ counterparts, resulting in an unequal density of markers. Thus, interpretation of ROHs in imputed ancient *Sus* should be taken with caution as it can be a result of the increase in homozygosity for Impute5. Overall, these downstream analyses highlight that there are biases and limitations toward imputation of *Sus scrofa* aDNA.

4.2 Factors Limiting the Power of Imputation of *Sus scrofa* aDNA

One of the limitations is size of the reference panel (59 individuals), but more specifically diversity in the reference panel. Studies on both humans and pigs showed that a larger and more diverse reference panel increase imputation accuracy (Jostins et al., 2011; Pistis et al., 2015; Van Den Berg et al., 2019; Ausmees et al., 2021; Wang et al., 2021). Ancient human imputation studies had a minimum of ~250 individuals to perform successful imputation (Ausmees, 2019). Adding individuals to the reference panel, that do not add genetic diversity to target samples does not increase genotype concordance, as seen from the results when adding Asian samples to the reference panel. The current reference panel lacks diversity, as the main groups in the reference panel consisted of European wild boar, European domestic pigs and Near Eastern wild boar, with (Dutch) European wild boar and European domestic pig dominating. A study on ancient human imputation observed a lower genotype concordance and similarity in their PCA for hunter-gatherer genomes of which ancestry is more or less absent in the reference panel (Ausmees et al., 2021). This was also found in imputation of pig breeds where rarer pig breeds had a lower genotype concordance and dosage score than breeds that were common in the reference panel (Wang et al., 2021). In this study genotype concordance improved when adding five ancient samples with Near Eastern ancestry to the reference panel. Improving and mitigating current biases of the reference panel should aid imputation. This could be achieved by including Mesolithic wild boar, Iberian, British, Scandinavian and East European wild boar, Near Eastern wild boar and domestic pigs to the reference panel.

Another potential limitation that is associated with the reference panel is the available reference genome. The *Sus scrofa* 11.1 reference genome, is from a Duroc individual, with known Asian introgression. Moreover, the nature of the reference genome could potentially increase the rate of false genotyping leading to errors in haplotypes and LD structure, which could result in decreasing imputation accuracy.

One final potential limitation is the genetic architecture of pigs. Accuracy of imputation is dependent on LD, recombination, genetic distance, and MAF (Stephens and Scheet, 2005; Browning et al., 2018; Ye et al., 2019). These factors are different in pigs compared to humans and even other livestock species, where average heterozygosity is lower and, LD and genetic distance are significantly greater (Zhang and Plastow, 2011). The recombination rate used in this study was based on nine breeding lines, all having introgression with Asian domestic pigs. This potentially introduced inaccuracies but is mitigated as the recombination was divided into bins of 1 MB, which might not be at a size resolution to introduce inaccuracies between wild and domestic pigs.

5 CONCLUSION

The use of imputation of ancient low-coverage *Sus scrofa* data resulted in relatively high genotype concordance and a moderate increase in information content. However, the imputed genotypes represented only a fraction, roughly 5%–50%, of all genotypes called in the HQ ancient genomes and featured biases toward the main population components in the reference panel. Our analysis indicated that these can lead to misidentifications or overrepresentation of ancestry components and selective traits in imputed genotypes. This is especially significant considering the weight archaeological debates place on ancestral relationships and admixture patterns of domesticated animals to understand the mechanisms of emergence and dispersal of early animal husbandry throughout the Neolithic across Europe and the Near East. This further highlights the measures needed to interpret the results and biases introduced by imputation and difficulty of imputation of admixed individuals. A more diverse reference panel is one of the most important priorities in ancient *Sus* imputation and particularly, introducing diversity present in ancient *Sus* could elevate accuracy and limit bias.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://www.ebi.ac.uk/ena/browser/view/PRJEB30282?show=reads Accession number PRJEB30282.

AUTHOR CONTRIBUTIONS

JE, DB, and OM conceptualized and designed the study. JE collected and assembled the data. JE, DB, and OM analyzed

the data. JE wrote the manuscript and JE, CÇ, OM, DB, and DR contributed to reviewing and editing the manuscript. DR and CÇ contributed to project oversight.

FUNDING

This study is supported by the Dutch Research Council Open Competition (Grant No. 406.18.HW.026) and the European Research Council award to DB under the European Union's Horizon 2020 research and innovation programme (885729-AncestralWeave).

ACKNOWLEDGMENTS

The authors would like to thank the Smurfit Institute of Genetics and the EDAN project for their helpful advice during discussions. The authors would also like to thank the reviewers for their comments and efforts towards improving our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872486/full#supplementary-material

REFERENCES

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

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

Ausmees, K. (2019). *Efficient Computational Methods for Applications in Genomics*. Dissertation. Uppsala: University of Uppsala.

Ausmees, K., Sanchez-Quinto, F., Jakobsson, M., and Nettelblad, C. (2021). An Empirical Evaluation of Genotype Imputation of Ancient DNA. BioRxiv [Preprint]. Available at https://www.biorxiv.org/content/10.1101/2021.12.22.473913v2.full.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinforma.* 43, 1110. doi:10.1002/0471250953.bi1110s43

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2009). Removal of Deaminated Cytosines and Detection of *In Vivo* Methylation in Ancient DNA. *Nucleic Acids Res.* 38 (6), e87. doi:10.1093/nar/gkp1163

Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J., et al. (2007). Novel High-Resolution Characterization of Ancient DNA Reveals C > U-type Base Modification Events as the Sole Cause of Post Mortem Miscoding Lesions. *Nucleic Acids Res.* 35 (17), 5717–5728. doi:10.1093/nar/gkm588

Browning, B. L., and Browning, S. R. (2013). Detecting Identity by Descent and Estimating Genotype Error Rates in Sequence Data. *Am. J. Hum. Genet.* 93 (5), 840–851. doi:10.1016/j.ajhg.2013.09.014

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015

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, 1084–1097. doi:10.1086/521987

Brunson, K., and Reich, D. (2019). The Promise of Paleogenomics beyond Our Own Species. *Trends Genet.* 35 (5), 319–329. doi:10.1016/j.tig.2019.02.006

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinforma.* 10 (421). doi:10.1186/1471-2105-10-421

Frantz, L. A. F., Haile, J., Lin, A. T., Scheu, A., Geörg, C., Benecke, N., et al. (2019). Ancient Pigs Reveal a Near-Complete Genomic Turnover Following Their Introduction to Europe. *Proc. Natl. Acad. Sci. U. S. A.* 116 (35), 17231–17238. doi:10.1073/pnas.1901169116

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., et al. (2014). Genome Flux and Stasis in a Five Millennium Transect of European Prehistory. *Nat. Commun.* 5 (5257). doi:10.1038/ncomms6257

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., and Orlando, L. (2011). mapDamage: Testing for Damage Patterns in Ancient DNA Sequences. *Bioinforma. Appl. Note* 27 (15), 2153–2155. doi:10.1093/bioinformatics/btr347

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of Pig Genomes Provide Insight into Porcine Demography and Evolution. *Nature* 491, 393–398. doi:10.1038/nature11622

Haldane's Mapping Function (2008). in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* (Dordrecht: Springer). doi:10.1007/978-1-4020-6754-9_7297

Hoss, M., Jaruga, P., Zastawny, T. H., Dizdaroglu, M., and Paabo, S. (1996). DNA Damage and DNA Sequence Retrieval from Ancient Tissues. *Nucleic Acids Res.* 24 (7), 1304–1307. doi:10.1093/nar/24.7.1304

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., and Kivisild, T. (202018542). Evaluating Genotype Imputation Pipeline for Ultra-low Coverage Ancient Genomes. *Sci. Rep.* 10. doi:10.1038/s41598-020-75387-w

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9 (3), 90–95. doi:10.1109/MCSE.2007.55

Johnsson, M., Whalen, A., Ros-Freixedes, R., Gorjanc, G., Chen, C.-Y., Herring, W. O., et al. (2021). Genetic Variation in Recombination Rate in the Pig. *Genet. Sel. Evol.* 53 (54). doi:10.1186/s12711-021-00643-0

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., and Orlando, L. (2013). mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters. *Bioinforma. Appl. Note* 29, 1682–1684.

Jostins, L., Morley, K. I., and Barrett, J. C. (2011). Imputation of Low-Frequency Variants Using the HapMap3 Benefits from Large, Diverse Reference Sets. *Eur. J. Hum. Genet.* 19 (6), 662–666. doi:10.1038/ejhg.2011.10

Kircher, M. (2012). "Analysis of High-Throughput Ancient DNA Sequencing Data," in *Analysis of High-Throughput Ancient DNA Sequencing Data" in Ancient DNA: Methods and Protocols, Methods in Molecular Biology*. Editors B. Shapiro, and M. Hofreiter (Springer Science + Business Media), 840, 197–228. doi:10.1007/978-1-61779-516-9_23

Kistler, L., Ware, R., Smith, O., Collins, M., and Allaby, R. G. (2017). A New Model for Ancient DNA Decay Based on Paleogenomic Meta-Analysis. *Nucleic Acids Res.* 45 (11), 6310–6320. doi:10.1093/nar/gkx361

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., et al. (2007). Ancient DNA, Pig Domestication, and the Spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. U.S.A.* 104 (39), 15276–15281. doi:10.1073/pnas.0703411104

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, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

MacHugh, D. E., Larson, G., and Orlando, L. (2017). Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annu. Rev. Anim. Biosci.* 5, 329–351. doi:10.1146/annurev-animal-022516-022747

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* 17 (1), 10–12. doi:10.14806/ej.17.1.200

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., et al. (2017). The Population Genomics of Archaeological Transition in West Iberia: Investigation of Ancient Substructure Using Imputation and Haplotype-Based Methods. *PLoS Genet.* 13 (7), e1006852. doi:10.1371/journal.pgen.1006852

McHugo, G. P., Dover, M. J., and MacHugh, D. E. (2019). Unlocking the Origins and Biology of Domestic Animals Using Ancient DNA and Paleogenomics. *BMC Biol.* 17 (98). doi:10.1186/s12915-019-0724-7

Meyer, M., and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* 2010 (6), pdb.prot5448. doi:10.1101/pdb.prot5448

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2015). Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* 32 (2), btv566–294. doi:10.1093/bioinformatics/btv566

Ottoni, C., Girdland Flink, L., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., et al. (2013). Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. *Mol. Biol. Evol.* 30 (4), 824–832. doi:10.1093/molbev/mss261

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., et al. (2004). Genetic Analyses from Ancient DNA. *Annu. Rev. Genet.* 38, 645–679. doi:10.1146/annurev.genet.37.110801.143214

Parks, M., and Lambert, D. (2015). Impacts of Low Coverage Depths and Post-mortem DNA Damage on Variant Calling: a Simulation Study. *BMC Genomics* 16 (9). doi:10.1186/s12864-015-1219-8

Pistis, G., Porcu, E., Vrieze, S. I., Sidore, C., Steri, M., Danjou, F., et al. (2015). Rare Variant Genotype Imputation with Thousands of Study-specific Whole-Genome Sequences: Implications for Cost-Effective Study Designs. *Eur. J. Hum. Genet.* 23 (7), 975–983. doi:10.1038/ejhg.2014.216

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., and Green, R. E. (2010). Computational Challenges in the Analysis of Ancient DNA. *Genome Biol.* 11 (R47), R47. doi:10.1186/gb-2010-11-5-r47

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

Ramírez, O., Burgos-Paz, W., Casas, E., Ballester, M., Bianco, E., Olalde, I., et al. (2015). Genome Data from a Sixteenth Century Pig Illuminate Modern Breed Relationships. *Hered. (Edinb)* 114, 175–184. doi:10.1038/hdy.2014.81

Rubinacci, S., Delaneau, O., and Marchini, J. (2020). Genotype Imputation Using the Positional Burrows Wheeler Transform. *PLoS Genet.* 16, e1009049. doi:10.1371/journal.pgen.1009049

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., and Delaneau, O. (2021). Efficient Phasing and Imputation of Low-Coverage Sequencing Data Using Large Reference Panels. *Nat. Genet.* 53, 120–126. doi:10.1038/s41588-020-00756-0

Sánchez-Quinto, F., Schroeder, H., Ramirez, O., Ávila-Arcos, M. C., Pybus, M., Olalde, I., et al. (2012). Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr. Biol.* 22 (16), 1494–1499. doi:10.1016/j.cub.2012.06.005

Slatkin, M., and Racimo, F. (2016). Ancient DNA and Human History. *Proc. Natl. Acad. Sci. U.S.A.* 113 (23), 6380–6387. doi:10.1073/pnas.1524306113

Song, H., Ye, S., Jiang, Y., Zhang, Z., Zhang, Q., and Ding, X. (2019). Using Imputation-Based Whole-Genome Sequencing Data to Improve the Accuracy of Genomic Prediction for Combined Populations in Pigs. *Genet. Sel. Evol.* 51 (58). doi:10.1186/s12711-019-0500-8

Stephens, M., and Scheet, P. (2005). Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. *Am. J. Hum. Genet.* 76 (3), 449–462. doi:10.1086/428594

Van den Berg, S., VandenPlas, J., van Eeuwijk, F. A., Bouwman, A. C., Lopes, M. S., and Veerkamp, R. F. (2019). Imputation to Whole-Genome Sequence Using Multiple Pig Populations and its Use in Genome-wide Association Studies. *Genet. Sel. Evol.* 51 (2). doi:10.1186/s12711-019-0445-y

Wang, Z., Zhang, Z., Chen, Z., Sun, J., Cao, C., Wu, F., et al. (2021). PHARP: A Pig Haplotype Reference Panel for Genotype Imputation. bioRxiv [Preprint]. doi:10.1101/2021.06.03.446888

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., et al. (2020). An Improved Pig Reference Genome Sequence to Enable Pig Genetics and Genomics Research. *Gigascience* 9 (6), giaa051. doi:10.1093/gigascience/giaa051

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R. P. M. A., Zinovieva, N., et al. (2017). Genome-wide SNP Data Unveils the Globalization of Domesticated Pigs. *Genet. Sel. Evol.* 49 (71). doi:10.1186/s12711-017-0345-y

Ye, S., Gao, N., Zheng, R., Chen, Z., Teng, J., Yuan, X., et al. (2019). Strategies for Obtaining and Pruning Imputed Whole-Genome Sequence Data for Genomic Prediction. *Front. Genet.* 10, 673. doi:10.3389/fgene.2019.00673

Zhang, C., and Plastow, G. (2011). Genomic Diversity in Pig (*Sus scrofa*) and its Comparison with Human and Other Livestock. *Cg* 12 (2), 138–146. doi:10.2174/138920211795564386

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

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

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fgene.2022.1008649

OPEN ACCESS

EDITED BY
Chunfang Zhao,
Anhui Science and Technology University, China

REVIEWED BY
Ran Di,
Chinese Academy of Agricultural Sciences, China
Zhenhui Li,
South China Agricultural University, China

*CORRESPONDENCE
Jing-Ting Shu,
shujingting@163.com

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

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

RECEIVED 01 August 2022
ACCEPTED 01 September 2022
PUBLISHED 15 September 2022

CITATION
Liu Y-F, Zhang M, Shan Y-J, Pang L-C, Ji G-G, Ju X-J, Tu Y-J, Shi S-Y, Bai H, Zou J-M and Shu J-T (2022), Transcriptome sequencing analysis of the role of miR-499-5p and SOX6 in chicken skeletal myofiber specification. *Front. Genet.* 13:1008649. doi: 10.3389/fgene.2022.1008649

Transcriptome sequencing analysis of the role of miR-499-5p and *SOX6* in chicken skeletal myofiber specification

Yi-Fan Liu[1,2†], Ming Zhang[1,2†], Yan-Ju Shan[1,2], Li-Chuan Pang[1,2], Gai-Ge Ji[1,2], Xiao-Jun Ju[1,2], Yun-Jie Tu[1,2], Shi-Ying Shi[1,2], Hao Bai[3], Jian-Min Zou[1,2] and Jing-Ting Shu[1,2*]

[1]Jiangsu Institute of Poultry Science Innovation Co., Yangzhou, China, [2]Key Laboratory for Poultry Genetics and Breeding of Jiangsu Province, Jiangsu Institute of Poultry Science, Yangzhou, China, [3]Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of China, Yangzhou University, Yangzhou, China

MicroRNAs (miRNAs) might play critical roles in skeletal myofiber specification. In a previous study, we found that chicken miR-499-5p is specifically expressed in slow-twitch muscle and that its potential target gene is *SOX6*. In this study, we performed RNA sequencing to investigate the effects of *SOX6* and miR-499-5p on the modulation and regulation of chicken muscle fiber type and its regulatory mechanism. The expression levels of miR-499-5p and *SOX6* demonstrated opposing trends in different skeletal muscles and were associated with muscle fiber type composition. Differential expression analysis revealed that miR-499-5p overexpression led to significant changes in the expression of 297 genes in chicken primary myoblasts (CPMs). Myofiber type-related genes, including *MYH7B* and *CSRP3*, showed expression patterns similar to those in slow-twitch muscle. According to functional enrichment analysis, differentially expressed genes were mostly associated with muscle development and muscle fiber-related processes. *SOX6* was identified as the target gene of miR-499-5p in CPM using target gene mining and luciferase reporter assays. *SOX6* knockdown resulted in upregulation of the slow myosin genes and downregulation of fast myosin genes. Furthermore, protein-protein interaction network analysis revealed that *MYH7B* and *RUNX2* may be the direct targets of *SOX6*. These results indicated that chicken miR-499-5p may promote slow-twitch muscle fiber formation by repressing *SOX6* expression. Our study provides a dataset that can be used as a reference for animal meat quality and human muscle disease studies.

KEYWORDS

chicken, meat quality, muscle fiber, RNA sequencing, miR-499-5p, Sox6

Introduction

Meat quality is an important livestock economic characteristic and is determined by muscle color, water holding capacity, pH, tenderness, flavor substance content, and intermuscular fat. As an essential tissue accounting for 40% of an animal's body weight, skeletal muscle contains muscle fibers that demonstrate diverse biochemical and structural features, and different fiber compositions in skeletal muscle are closely associated with muscle quality (Lee et al., 2010; Joo et al., 2013; Ismail and Joo, 2017). Myofibers within chickens are classified as white and red fibers; they indicate glycolytic (type IIB; undergo glycolytic metabolism and rapid contraction) and oxidative (type I/IIA; undergo oxidative metabolism and slow contraction) fibers, respectively (Zierath and Hawley, 2004; Hakamata et al., 2018).

The skeletal muscle fiber type is directly determined by various genetic factors. miRNAs are small (22 nucleotides) non-coding RNAs, which show negative gene regulation at the post-transcription level by combining with the 3'-untranslated region (3'UTR) in target mRNAs (Gladka et al., 2012). Several studies have suggested that miRNAs regulate skeletal muscle function, including fiber-type transition and differentiation (Zhang et al., 2014; Zhang et al., 2019; Yin et al., 2020).

miR-499-5p has been identified as a myosin-encoding miRNA, which is recognized as an essential regulatory factor for muscle fiber conversion and as a muscle disease biomarker (Liu et al., 2016; Wan et al., 2018; Shi et al., 2019). Van Rooij et al. first reported that miR-499-5p and miR-208b knockouts in mice resulted in slow-to-fast alteration of muscle fiber type (van Rooij et al., 2009). In mammals and fish, miR-499 is highly expressed within slow-twitch muscles, which possibly modulates the types of muscle fibers via target mRNA, including *SOX6*, *THRAP1*, *FNIP1*, *ROD1,* and pathways, including NFATc1/MEF2C and FNIP1/AMPK (Nachtigall et al., 2015; Liu et al., 2016; Wang et al., 2017; Xu et al., 2018). Nonetheless, the possible role of miR-499-5p and related mechanisms in chicken muscle fiber switching have not been clarified.

RNA-seq is an effective tool for gene function studies because it can obtain global gene expression levels in tissues or cells with high accuracy (Liu et al., 2017; Pu et al., 2022). In this study, RNA-seq was used to explore the biological function of miR-499-5p in chicken skeletal muscle fiber specification and to identify the target gene of miR-499-5p in chicken primary myoblasts (CPMs). Our findings may enable a better understanding of the underlying mechanisms implicated in the quality performance of chicken meat.

Materials and methods

Muscle sampling

The Wenchang chickens used for ATPase staining and quantitative real-time PCR (qRT-PCR) assays were obtained from Tanniu Chicken Co., Ltd. (Haikou, China). Six 20-weeks-old hens of similar body weights (1,315–1,430 g) were slaughtered by electrical stunning and exsanguination. Intermediate section samples of various skeletal muscles were collected after rapid dissection, followed by freezing in liquid nitrogen and preservation at -80°C. Skeletal muscle samples included the *pectoralis major* (PM), *sartorius* (SA), *lateral pectineus* (LP), and *medial gastrocnemius* (MG) muscles.

Myosin ATPase staining

According to our previous report, myofiber types were identified using myosin ATPase staining (Li et al., 2016). As described by Brooke and Kaiser (1970), type IIB was intensely active and stained dark brown, type IIA was weakly active and stained light brown, and type was inactive and non-stained.

Cell culture and transfection

We separated CPMs in E11 chicken leg muscles according to a previous description, followed by culturing in DMEM/F12 medium (Gibco, Carlsbad, CA, United States) containing 20% fetal bovine serum (FBS; Gibco) (Shan et al., 2020). When the density reached 80%, FBS was replaced with 5% horse serum (Procell, Wuhan, China) to induce myogenic differentiation. miR-499-5p mimics, NC inhibitor, siRNA targeting *SOX6*, or siRNA negative control were transfected into CPMs using Lipofectamine 3,000 (Invitrogen, Carlsbad, CA, United States) following the manufacturer's protocol. The miRNA mimics and siRNAs were prepared by Ribobio Co., Ltd (Guangzhou, China). The *SOX6* siRNA sequence was 5'-GGTGAATACAAACAACTG A-3'. After 48 h of transfection, each cell group with four replicates was collected for the subsequent RNA-seq and qRT-PCR assays.

RNA extraction and sequencing

Total RNA was extracted from CPMs using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) following the manufacturer's protocol. An Agilent 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States) and a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States) were used to check RNA quality. High-throughput RNA-seq was performed on an Illumina HiSeq 4,000 platform (Illumina, San Diego, CA, United States) at Gene Denovo Biotechnology Co. (Guangzhou, China).

High-throughput sequencing data analysis

For quality control, fastp v0.18 with default parameters was utilized to process the raw data, followed by alignment of paired-end clean reads against the reference genome (*GRCg6a*) with Hisat2 (Sirén et al., 2014; Chen et al., 2018). FeatureCounts were used to count reads mapped into diverse genes, and fragments per kilobase million (FPKM) values were determined for diverse genes (Liao et al., 2014). Differentially expressed genes (DEGs) in the treatment and control groups were identified using the DESeq2 R package (Love et al., 2014). DEGs were identified based on the threshold of false discovery rate (FDR) < 0.05 with DESeq2. Target genes of miR-499-5p were identified using the TargetScan 7.2 (http://www.targetscan.org/vert_72/) and miRDB (http://mirdb.org/) web platforms with default parameters.

Gene ontology (GO) enrichment analysis of DEGs was performed using an online tool from OMICSHARE (www.omicshare.com/tools). Significant enrichment of GO terms was indicated by p-value < 0.05. OMICSHARE was used to identify significantly enriched DEGs in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The STRING database (https://string-db.org/) was utilized for protein-protein interaction (PPI) network construction at moderate confidence. Cytoscape software v3.5.1 was used for visualization.

qRT-PCR

The Primer3 website (https://bioinfo.ut.ee/primer3-0.4.0/) and miRprimer2 software (https://sourceforge.net/projects/mirprimer) were used to design qRT-PCR primers for genes and miR-499-5p, respectively. Primer information is listed in Supplementary Table S1. For mRNA quantification, cDNA was prepared using the HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper) (R312, Vazyme, Nanjing, China). qRT-PCR with ChamQ SYBR Color qRT-PCR Master Mix (Q411, Vazyme) was performed for cDNA quantification using a real-time PCR system (Mx3000P, Agilent Technologies). For miRNA quantification, cDNA was synthesized using the miRcute Plus miRNA First-Strand cDNA kit (KR211, Tiangen, Beijing, China), and qRT-PCR was performed using miRcute Plus miRNA qPCR kit (SYBR Green) (FP411, Tiangen). Chicken *GAPDH* and U6 snRNA were used as endogenous references. The $2^{-\Delta\Delta Ct}$ method was used to calculate the fold-change.

Luciferase reporter assay

DF-1 cells were used for the luciferase reporter assay. The PmiR-RB-Report vector was used to synthesize mutant (MUT) and wildtype (WT) luciferase reporter vectors at Ribobio Co., Ltd (Guangzhou, China). The information on mutant sites in the luciferase reporter vectors is presented in Supplementary Table S2**.** The cells were transfected with MUT or WT constructs and miR-499-5p mimics or NC mimics. After 48 h, luminescence was detected using a Dual-Glo Luciferase Assay System (E2920, Promega, Madison, WI, United States).

Western blotting

CPMs were seeded in six-well plates and transfected with overexpression miRNA mimics for 48 h. Cells were harvested, washed with 1× phosphate buffered saline (PBS) and lysed in RIPA lysis buffer. Immunoblotting was performed using standard procedures. The primary antibodies used were anti-SOX6 (NBP2-20458, NOVUS, Saint Louis, MI, United States) and anti-GAPDH (ERAB0003, Erwanbiotech, Shanghai, China). The goat anti-rabbit IgG-HRP (111-035-003; Jackson immune Research, Bar Harbor, ME, United States) was used as the secondary antibody.

Statistical analysis

The results of the fiber type content, miR-499-5p, and gene levels within muscles and cells are presented as mean ± SD. The expression patterns of the genes and miR-499-5p were statistically compared among the different muscle groups using the one-way analysis of variance (ANOVA) followed by Tukey's test, as implemented in the SPSS Software Suite (SPSS version 20.0). The statistical significance of qRT-PCR analysis results and luciferase reporter assay results between the treatment and control groups was determined using t-test. Differences were considered statistically significant at p-values of < 0.05 and < 0.01.

Results

miR-499-5p and *SOX6* expression levels in different skeletal muscles

As shown in Figures 1A,B, four different skeletal muscles—the PM, SA, LP, and MG, exhibit different muscle fiber compositions. PM is composed entirely of type IIB fiber, whereas the other three contain three types of muscle fibers. SA has the highest content of type IIA fibers, whereas MG has the most amount of type fiber.

The qRT-PCR results showed that miR-499-5p and *SOX6* have opposing expression patterns. As shown in Figure 1C, miR-499-5p showed the highest expression in MG, followed

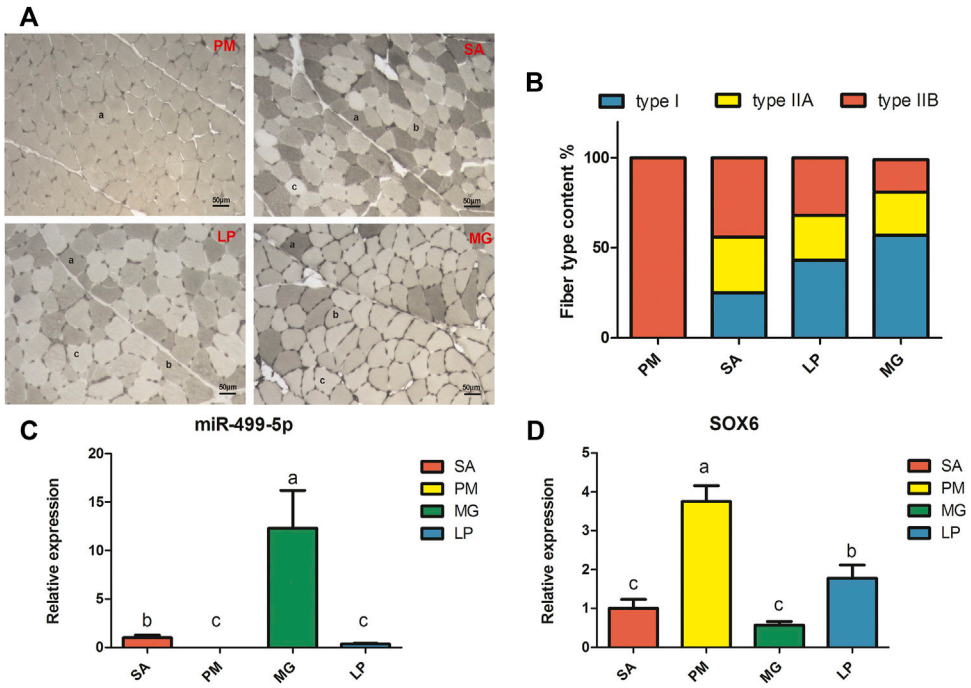


FIGURE 1
Myofiber type phenotype and gene expression levels in four different skeletal muscles. **(A)** Myosin ATPase staining of four different skeletal muscles. PM: *pectoralis major*, SA: *sartorius*, LP: *lateral pectineus*, MG: *medial gastrocnemius*. **(A)** type IIB fibers, **(B)** type IIA fibers, **(C)** type I fibers. **(B)** Fiber type content of four different skeletal muscles. Relative expression levels of miR-499-5p **(C)** and *SOX6* **(D)** in four different skeletal muscles. All results are shown as mean ± SD. $p < 0.05$ or less is considered significant and indicated with different letters.

by SA and LP, and the lowest expression in PM. As expected, *SOX6* was most significantly upregulated in PM, followed by LP, SA, and MG (Figure 1D). The expression level of miR-499-5p was found to have a significant positive correlation with type I fiber content and negative correlation with type IIB fiber content in skeletal muscles, whereas *SOX6* expression showed a similar correlation with type IIB fiber content (Supplementary Table S3).

Effect of miR-499-5p overexpression on CPM transcriptome

After 48 h of miR-499-5p overexpression in the CPMs, the qRT-PCR results showed a highly significant increase in miR-499-5p expression compared with that in the control ($p < 0.01$, Figure 2). Four miR-499-5p overexpression and four control CPMs were collected for transcriptome sequencing. A total of 297 DEGs were detected (FDR <0.05), including 127 upregulated and 170 downregulated genes within cells overexpressing miR-499-5p compared with control cells (Supplementary Table S4). GO enrichment analysis showed that DEGs were mainly related to muscle development, muscle contraction, and muscle cell differentiation (Figure 3A). KEGG pathway analysis showed that the enriched pathways of DEGs included many muscle function-related pathways such as actin cytoskeleton regulation, focal adhesion, and cardiac muscle contraction (Figure 3B).

Furthermore, several key genes associated with muscle fiber type were found to be differentially expressed in miR-499-5p overexpressing CPMs (Table 1). Compared with our previous transcriptome data on slow- and fast-twitch chicken muscles, most of these genes shared similar expression patterns in miR-499-5p-overexpressing CPMs and those in slow-twitch muscles (Liu et al., 2020). A series of myosin genes, including *MYH1B*, *MYH1E*, and *MYH7B*, were upregulated after miR-499-5p overexpression. *MYH7B*, which encode slow myofibrillar proteins, was highly expressed in the overexpressed cells (log2foldchange = 2.42). The *MYH1B*, *MYH1E*, and *MYH7B* expression patterns were confirmed using qRT-PCR (Figure 2).

Target mining of miR-499-5p using transcriptome data

We used TargetScan and mirDB for miR-499-5p target gene prediction and obtained the target gene information with

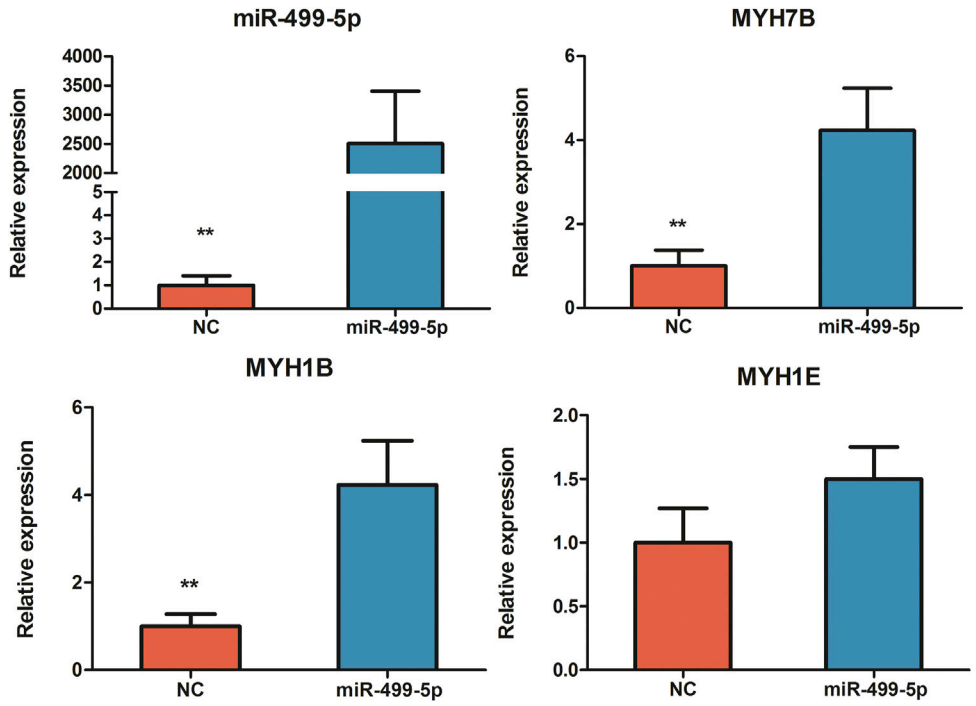


FIGURE 2
The relative expression levels of miR-499-5p, *MYH7B*, *MYH1B*, and *MYH1E* in miR-499-5p overexpressing CPMs and negative control by qRT-PCR. All results are shown as mean ± SD. $^{**}p < 0.01$ as compared with the control.

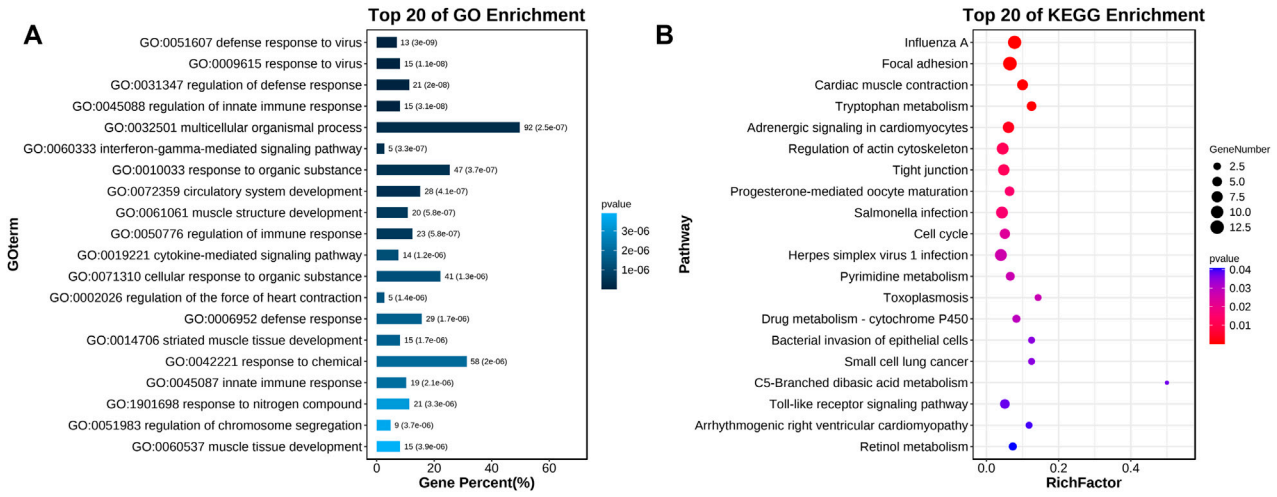


FIGURE 3
Functional enrichment of DEGs between miR-499-5p overexpressing and NC CPMs. **(A)** Top 20 significantly enriched GO-BP terms; **(B)** Top 20 significantly enriched KEGG pathways.

high confidence by comparing it with transcriptome data. We first identified 444 target genes of miRNA-499-5p in CPMs that could be identified as target genes using at least one tool (Supplementary Table S5). We found that the expression of more than 65% of the target genes had reduced after miRNA overexpression, with 16 reaching

TABLE 1 Expression change of myofiber type-related DEGs between miR-499-5p overexpressing and NC CPMs. Trends in gene expression changes between the present study and our previous study on slow-twitch and fast-twitch muscles in chickens are also shown (Liu et al., 2020).

Gene symbol	log_2(FC) in miR-499-5p vs. NC CPM	log_2(FC) in slow vs. fast muscles	FDR	Expression change trends between studies
AKAP6	−0.37	2.08	5.7E-06	Inconsistent
ATP2A2	0.50	6.14	3.85E-43	Consistent
CAMKK2	−0.49	1.35	0.0002	Inconsistent
CAPN3	−0.56	−2.30	7.7E-07	Consistent
CAV1	0.28	1.75	3.97E-06	Consistent
CAV3	−0.27	−2.25	0.0001	Consistent
CSRP3	−0.34	−1.40	1.68E-09	Consistent
FGF10	0.28	1.01	5.83E-06	Consistent
FILIP1	−0.33	−10.46	3.97E-06	Consistent
ITGA4	0.26	1.09	0.0012	Consistent
KLHL40	−0.35	−1.20	3.16E-07	Consistent
KLHL41	−0.53	1.10	1.05E-37	Inconsistent
MYBPC1	0.68	10.50	5.97E-26	Consistent
MYH1A	1.09	5.36	1.12E-27	Consistent
MYH1B	1.29	5.18	2.1E-104	Consistent
MYH1C	0.85	2.12	5.42E-31	Consistent
MYH1E	0.62	−7.87	1.07E-11	Inconsistent
MYH1F	0.79	1.81	6.87E-21	Consistent
MYH1G	1.08	7.13	2.13E-48	Consistent
MYH7B	2.42	5.25	1.29E-11	Consistent
MYL10	0.66	11.25	1.12E-07	Consistent
MYL2	1.30	7.59	0.0014	Consistent
MYL3	1.27	16.86	1.22E-88	Consistent
MYLK2	0.42	−1.37	0.0005	Inconsistent
MY O 7L2	0.48	6.90	0.0002	Consistent
MYOM2	−0.56	−2.07	3.02E-26	Consistent
PAK1	0.38	1.30	0.0185	Consistent
PITX1	−0.34	8.99	0.0089	Inconsistent
SOX6	−0.83	−6.71	0.0423	Consistent
TNNI1	0.63	9.41	3.09E-26	Consistent
TNNT1	0.30	9.71	7.38E-06	Consistent
TNNT2	0.29	1.32	1.85E-12	Consistent
TPM2	0.34	2.07	5.21E-10	Consistent

significant levels (FDR <0.05, Table 2). To improve the target gene prediction accuracy, we used the intersection of the results of the two tools as the possible targets of miR-499-5p. A total of 72 target genes were identified in CPMs, four of which were significantly downregulated in the overexpressed cells (FDR <0.05).

Among the four significantly downregulated target genes, *SOX6* showed the most significant decrease in expression (log2foldchange = -0.83). In addition, target genes associated with myofiber-type switching, including *PPP3CB*, *FNIP1*, and *FNIP2*, were identified, but they presented only lower expression changes.

Validation of targeted binding relationship between miR-499 and target genes

We used qRT-PCR, western blotting, and luciferase reporter assay to validate the target binding relationship between miR-499-5p and the target genes (Figure 4). The qRT-PCR results revealed that the relative *SOX6*, *PPP3CB*, and *FNIP1* expression levels decreased in the overexpressed cells, with *SOX6* showing the most remarkable change in expression, consistent with the RNA-seq results. Overexpression of miR-499-5p decreased the protein level of SOX6. The luciferase reporter assay showed that miR-499-5p significantly decreased luciferase activity in

TABLE 2 Expression change of miR-499-5p target genes in miR-499-5p overexpressing CPMs compared with negative control. TargetScan and mirDB were used for miR-499-5p target gene prediction.

Gene symbol	Vs. TargetScan	Vs. mirDB	log_2 (foldchange)	*P* -value	FDR
SOX6	yes	yes	−0.83	0.0017	0.0423
NRIP1	yes	yes	−0.29	3.72E-05	0.0032
OSBPL1A	yes	yes	−0.20	0.0016	0.0403
PPP3CB	yes	yes	−0.10	0.0094	0.0438
LARP4B	yes	no	−0.25	0.001	0.0438
FNIP1	yes	no	−0.17	0.0018	0.0456
OLFML1	no	yes	−1.11	2.98E-07	5.76E-05
PARP9	no	yes	−0.61	7.55E-06	0.0008
ZNF106	no	yes	−0.42	6.46E-15	5.28E-12
SAMD8	no	yes	−0.34	6.17E-06	0.0007
NKTR	no	yes	−0.27	0.0006	0.0292
ESF1	no	yes	−0.25	0.0014	0.0466
UNKL	no	yes	−0.24	0.0001	0.0102
PTN	no	yes	−0.21	0.0001	0.0093
EGLN1	no	yes	−0.19	8.67E-07	0.0001
LZIC	no	yes	−0.19	0.0001	0.0077
FNIP2	no	yes	−0.13	0.0077	0.0488

combination with *SOX6* 3'-UTR sites (p-value < 0.05). However, the *PPP3CB* and *FNIP1* luciferase activities were not significantly altered following miR-499-5p transfection (p-value > 0.05).

Effect of *SOX6* on muscle fiber specification

To investigate the effect of *SOX6* on chicken muscle fiber specification, RNA interference and RNA-seq techniques were used to obtain the expression patterns of muscle fiber-related genes and to construct a regulatory network for muscle fiber specification. After RNA interference, *SOX6* expression was 0.6- and 2.6-fold downregulated in CPMs, as determined using qRT-PCR and RNA-seq, respectively (Figure 5 and Table 3). We identified 1,448 DEGs in si-*SOX6* and control CPMs using FDR <0.05 as the screening criterion. Among these, 107 genes overlapped with the results of the miR-499-5p overexpression experiment (Supplementary Table S6). The foldchange of *MYH7B*, *MYH1B*, *CSRP3*, *TNNI1*, *MYL2* and *MYL10* expression were confirmed using qRT-PCR (Figure 6).

The myosin gene expression pattern revealed that the knockdown of *SOX6* resulted in significant upregulation of slow-twitch muscle-associated *MYH7*, *MYH7B*, and *MYH1B* and significant downregulation of fast-twitch muscle-associated *MYH1D* and *MYH1F* (Table 3). A total of 24 differentially expressed genes, known for myofiber type switching, were selected to build a protein-protein interaction network based on the STRING database. Among them, a total of six genes, *MYH7B*, *MYH1B*, *CSRP3*, *TNNI1*, *MYL2* and *MYL10,* were down-regulated after *SOX6* interference, while the remaining genes were up-regulated. According to the network, *SOX6* may have a direct regulatory relationship with *RUNX2*, *PRDM1*, *HDAC4*, and *MYH7B* (Figure 7).

Discussion

Improving meat quality is the ultimate goal of broiler breeding, especially in native Chinese breeds, and the composition of muscle fibers is a critical factor affecting meat quality. For instance, high oxidative fiber level facilitates meat flavor and juiciness (Sibut et al., 2011). Muscle fiber characteristics are affected by muscle type, location, and function within an animal (Shu et al., 2017). To reveal the regulatory mechanisms affecting myofiber type in chicken skeletal muscle, we analyzed miRNA expression profiles in oxidative and glycolytic muscles in our previous study (Liu et al., 2020). miR-499-5p showed specific expression in oxidative muscles, consistent with the results observed in fish and pigs, suggesting that this miRNA is a crucial factor in regulating chicken myofiber type conversion (Duran et al., 2015; Jiang et al., 2021). In this study, a comparison of different muscles revealed a positive correlation between miR-499-5p and type I muscle fiber content and a negative correlation between miR-499-5p and type IIB muscle fiber content, implying that miR-499-5p is associated with meat quality traits.

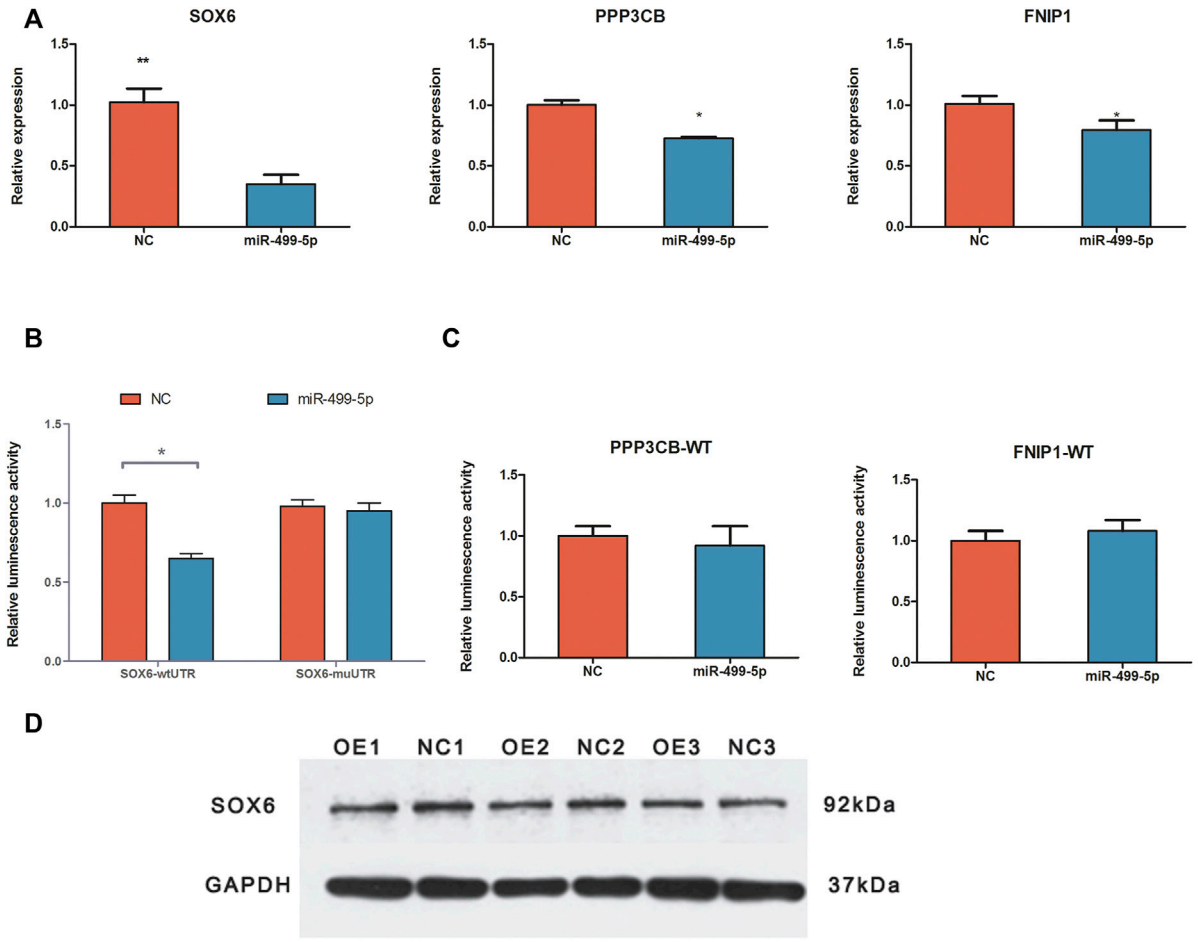


FIGURE 4
Validation of targeted binding relationship between miR-499-5p and target genes. **(A)** The relative expression levels of *SOX6*, *PPP3CB*, and *FNIP1* in miR-499-5p overexpressing CPMs and negative control by qRT-PCR; **(B)** Results of luciferase reporter assay for *SOX6* and miR-499-5p; miR-499-5p binding site mutant and wildtype vector of *SOX6*'s 3'UTR were used. **(C)** Results of the luciferase reporter as say for *PPP3CB*/miR-499-5p and *FNIP1*/miR-499-5p; only wildtype vectors were used. All results are shown as mean ± SD. $^{**}p < 0.01$ and $^{*}p < 0.05$ as compared with control. **(D)** The protein levels of *SOX6* in miR-499-5p overexpressing CPMs and negative control by western bolting.

miR-499-5p belongs to MyomiRs, which are mainly expressed in the heart and skeletal muscles and can be encoded via specific introns in slow muscle-specific myosin *MYH7B* (McCarthy, 2011). miR-499-5p positively regulates the formation of slow-twitch muscle fibers and negatively regulates the formation of fast-twitch muscle fibers in mouse and pig models (van Rooij et al., 2009; Wang et al., 2017). Although miRNAs are highly conserved among species, there is a lack of evidence regarding the effect of miR-499-5p on chicken myofiber type specification (Lee et al., 2007). In this study, RNA-seq was used to investigate the effect of miR-499-5p on the CPM transcriptome. Therefore, miR-499-5p overexpression markedly enhanced *MYH7B* and *MYH1B* mRNA expression, demonstrating its ability to promote slow-twitch muscle fiber formation in chickens. Many myofiber type-related genes were differentially expressed after overexpression and showed a similar expression pattern to that of the slow-twitch muscle. For example, cysteine and glycine-rich protein 3 (*CSRP3*), also known as muscle LIM protein (MLP), which promotes fast-twitch muscle fiber formation in chickens, showed significantly decreased mRNA expression post miR-499-5p overexpression (Shan et al., 2022). Calpain 3 (*CAPN3*) is the only muscle-specific calpain that has impotant role in muscle function such as muscle formation and remodeling (Chen et al., 2021). Studies have indicated that mutations in *CAPN3* are associated with meat quality traits including muscle fiber composion in chicken and pig (Gandolfi et al., 2011; Felício et al., 2013). In the present study, *CAPN3* was found to be downregulated following miR-499-5p overexpression, suggesting their roles in the regulation of muscle fiber types. Functional enrichment analysis revealed that DEGs are involved in several muscle development-related processes and signaling pathways

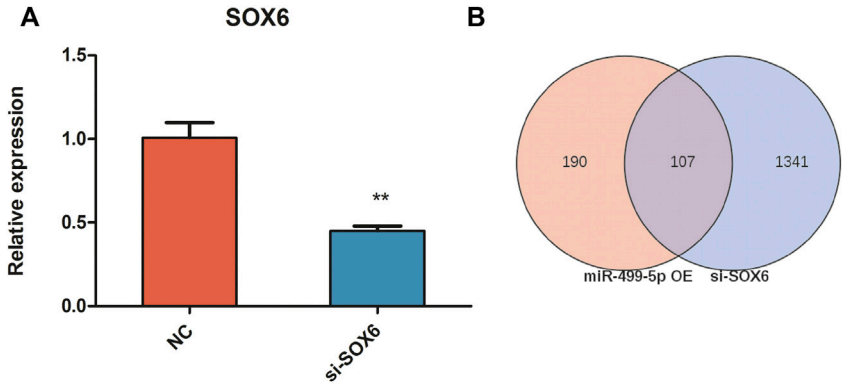


FIGURE 5
The inhibition efficiency of si-*SOX6*. **(A)** The relative expression levels of *SOX6* in *SOX6* knockdown CPMs and negative control by qRT-PCR. **(B)** The Venn diagram of differential genes in miR-499 overexpressing and *SOX6* knockdown assays. All results are shown as mean ± SD. $^{**}p < 0.01$ as compared with the control.

TABLE 3 Expression change of *SOX6* and myosin genes in *SOX6* knockdown CPMs compared with negative control.

Gene symbol	NC	si-*SOX6*	$\log_2$ (fc)	*P* -value	FDR
SOX6	3.36	1.27	−1.41	5.07E-07	3.08E-05
MYH1A	19.41	14.49	−0.42	0.0455	0.2441
MYH1B	198.12	329.25	0.73	0.0275	0.1227
MYH1D	1,123.47	509.12	−1.14	0.003	0.0234
MYH1E	50.04	27.38	−0.87	0.0383	0.1795
MYH1F	113.74	53.80	−1.08	0.04	0.1606
MYH7	0.50	0.75	0.59	0.0181	0.1536
MYH7B	0.69	1.25	0.85	0.043	0.1968
MYH15	835.19	422.29	−0.98	1.10E-09	2.58E-07

associated with myofiber-type switching, including focal adhesion and regulation of the actin cytoskeleton (Ohnuki et al., 2013; Graham et al., 2015). Collectively, miR-499-5p promotes slow-twitch muscle fiber formation through diverse regulatory factors and signaling pathways.

miRNAs are involved in regulating biological processes by binding to the 3'UTR of target mRNA to repress gene expression (Bartel, 2009). Transcriptome data and bioinformatic tools were used to identify miR-499-5p target genes. More than 60% of the predicted target genes showed decreased expression in the CPM, indicating that the miRNA transfection in this study was effective. Under stringent prediction criteria, *SOX6* was identified as the target gene with the highest decrease in expression. *SOX6* is a transcription factor that has a critical effect on embryonic muscle development and myofiber maintenance (Hagiwara et al., 2005; Hagiwara et al., 2007; Quiat et al., 2011; Jackson et al., 2015). In chickens, copy number variation in *SOX6* is associated with muscle development and body weight (Lin et al., 2018; Fernandes et al., 2021). Previous studies have reported that *SOX6* is a target of miR-499-5p in skeletal and cardiac muscles (Jia et al., 2016; Wang et al., 2017). Our results revealed opposing expression trends of miR-499-5p and *SOX6* in different chicken skeletal muscles, consistent with the results observed in Nile tilapia and pigs (Nachtigall et al., 2015; Wang et al., 2017). The luciferase reporter assay showed that *SOX6* is the direct target of miR-499-5p, which may regulate chicken myofiber specification by controlling *SOX6* expression. Other genes, including *FNIP1* and *THRAP1*, have also been reported to be important targets of miR-499-5p related to muscle fiber-type specification (Liu et al., 2016; Xu et al., 2018). *THRAP1* did not show a significant decrease in expression after miR-499-5p overexpression, and *FNIP1* was not significantly associated with miR-499-5p. This study revealed that *SOX6* is the primary target of miR-499-5p in chicken myofiber-type conversion.

To further understand the effect of *SOX6* on chicken myofiber type specification and the regulatory network involved, we performed a transcriptome analysis of *SOX6*-knockdown CPMs. As expected, the *SOX6* knockdown resulted in the formation of slow-twitch muscle fibers and

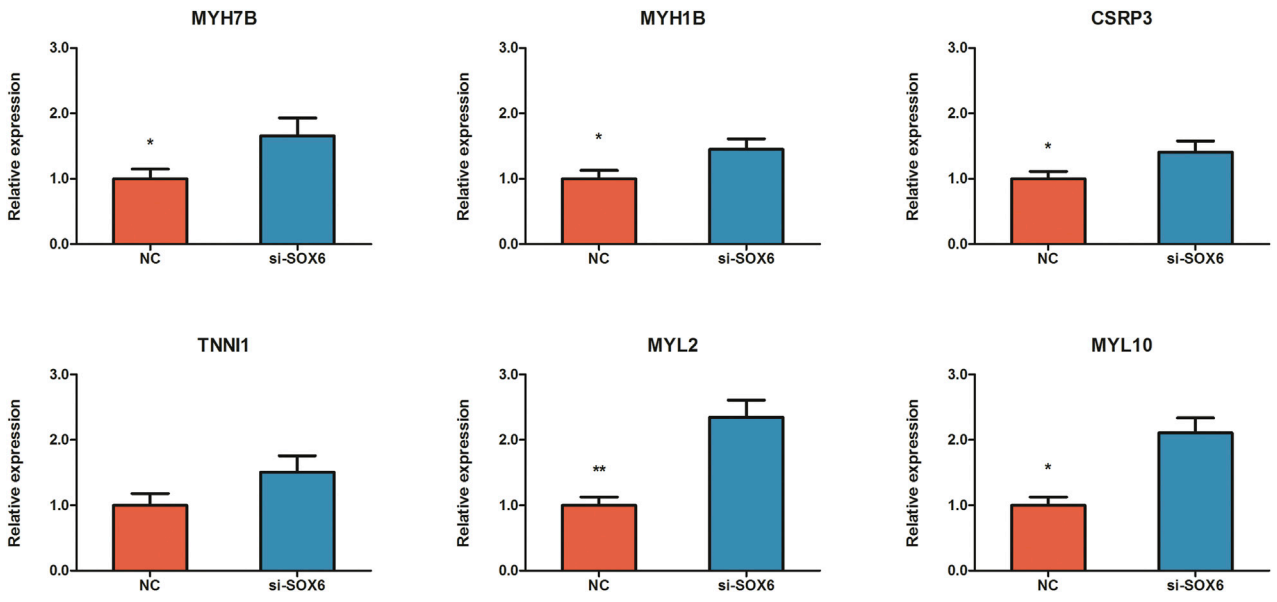


FIGURE 6
Validaton of expression change of *MYH7B*, *MYH1B*, *CSRP3*, *TNNI1*, *MYL2* and *MYL10* between *SOX6* knockdown and NC CPMs using qRT-PCR. All results are shown as mean ± SD. $^{**}p < 0.01$ and $^{*}p < 0.05$ as compared with control.

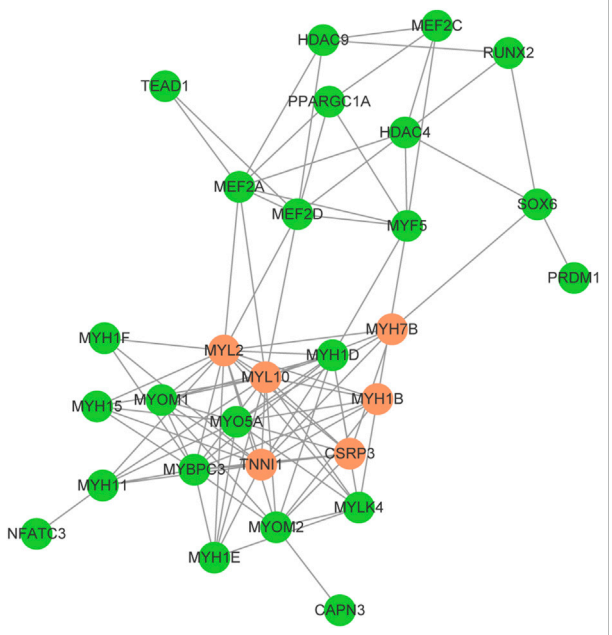


FIGURE 7
Protein-protein interaction network for the selected DEGs between *SOX6* knockdown and NC CPMs. Upregulated genes are shown in red and downregulated genes are shown in blue.

a decrease in fast-twitch muscle fibers. Protein-protein interaction network analysis revealed that *SOX6* could act directly on myosin and functional genes related to muscle fiber type. An et al. found that *SOX6* could directly target and bind *MYH7B* and inhibit its expression using CHIP-seq analysis, and the same result was obtained in the present study (An et al., 2011). Although fast-twitch muscle myosin genes, such as *MYH1D* and *MYH1F*, were significantly reduced after *SOX6* knockdown, there was no evidence that *SOX6* could act directly on them. *RUNX2* is a critical transcription factor, and *SOX6* and *RUNX2* interact during osteogenesis (Zhang et al., 2015). *RUNX2* has not been determined to play a role in the specification of muscle fiber types; combined with our findings, we hypothesize that *SOX6* and *RUNX2* can initiate muscle fiber type-related signaling pathways such as the PGC1α/MEF2C pathway (Shan et al., 2020).

In conclusion, we first investigated the role of miR-499-5p and *SOX6* in regulating myofiber type in chicken skeletal muscle using RNA-seq. The results revealed that both miR-499-5p and *SOX6* enhanced slow-twitch muscle fiber formation and that *SOX6* is the primary target of miR-499-5p in this process. This study provides detailed transcriptome sequencing results and serves as a reference for related studies on chicken meat quality and other species.

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://ngdc.cncb.ac.cn/gsa, CRA007272.

Ethics statement

The animal study was reviewed and approved by the Animal Care and Use Committee at the Jiangsu Institute of Poultry Science.

Author contributions

Y-FL, MZ, and J-TS contributed to conception and design of the study. Y-FL, Y-J-TS, L-CP, G-GJ, and X-JJ performed the experiments and participated in data collection. MZ, Y-JT, HB, and S-YS analyzed the data. Y-FL and MZ drafted the manuscript. J-MZ and J-TS revised the manuscript critically. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was funded by the Natural Science Foundation of Jiangsu Province (BK20211121 and BK20210955), the 'JBGS' Project of Seed Industry Revitalization in Jiangsu Province (JBGS [2021]107), the Special Fund for Independent Innovation of Agricultural Science and Technology in Jiangsu Province of China (CZ(20)3001), the Open Project Program of Joint International Research Laboratory of Agriculture and Agri-Product Safety, the Ministry of Education of China, Yangzhou University (JI-LARKF202020).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1008649/full#supplementary-material

References

An, C. I., Dong, Y., and Hagiwara, N. (2011). Genome-wide mapping of *SOX6* binding sites in skeletal muscle reveals both direct and indirect regulation of muscle terminal differentiation by *SOX6*. *BMC Dev. Biol.* 11, 59. doi:10.1186/1471-213X-11-59

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

Brooke, M. H., and Kaiser, K. K. (1970). Muscle fiber types: How many and what kind? *Arch. Neurol.* 23, 369–379. doi:10.1001/archneur.1970.00480280083010

Chen, L., Tang, F., Gao, H., Zhang, X., Li, X., and Xiao, D. (2021). CAPN3: A muscle-specific calpain with an important role in the pathogenesis of diseases (review). *Int. J. Mol. Med.* 48, 203. doi:10.3892/ijmm.2021.5036

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34, i884-i890. doi:10.1093/bioinformatics/bty560

Duran, B. O. d. S., Fernandez, G. J., Mareco, E. A., Moraes, L. N., Salomão, R. A. S., Gutierrez de Paula, T., et al. (2015). Differential microRNA expression in fast- and slow-twitch skeletal muscle of Piaractus mesopotamicus during growth. *PLOS ONE.* 10, e0141967. doi:10.1371/journal.pone.0141967

Felício, A. M., Boschiero, C., Balieiro, J. C., Ledur, M. C., Ferraz, J. B., Michelan Filho, T., et al. (2013). Identification and association of polymorphisms in CAPN1 and CAPN3 candidate genes related to performance and meat quality traits in chickens. *Genet. Mol. Res.* 12, 472–482. doi:10.4238/2013.February.8.12

Fernandes, A. C., da Silva, V. H., Goes, C. P., Moreira, G. C. M., Godoy, T. F., Ibelli, A. M. G., et al. (2021). Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics*. 22, 354. doi:10.1186/s12864-021-07676-1

Gandolfi, G., Pomponio, L., Ertbjerg, P., Karlsson, A. H., Nanni Costa, L., Lametsch, R., et al. (2011). Investigation on CAST, CAPN1 and CAPN3 porcine gene polymorphisms and expression in relation to post-mortem calpain activity in muscle and meat quality. *Meat Sci.* 88, 694–700. doi:10.1016/j.meatsci.2011.02.031

Gladka, M. M., da Costa Martins, P. A., and De Windt, L. J. (2012). Small changes can make a big difference—microRNA regulation of cardiac hypertrophy. *J. Mol. Cell. Cardiol.* 52, 74–82. doi:10.1016/j.yjmcc.2011.09.015

Graham, Z. A., Gallagher, P. M., and Cardozo, C. P. (2015). Focal adhesion kinase and its role in skeletal muscle. *J. Muscle Res. Cell. Motil.* 36, 305–315. doi:10.1007/s10974-015-9415-3

Hagiwara, N., Ma, B., and Ly, A. (2005). Slow and fast fiber isoform gene expression is systematically altered in skeletal muscle of the *SOX6* mutant, p100H. *Dev. Dyn.* 234, 301–311. doi:10.1002/dvdy.20535

Hagiwara, N., Yeh, M., and Liu, A. (2007). *SOX6* is required for normal fiber type differentiation of fetal skeletal muscle in mice. *Dev. Dyn.* 236, 2062–2076. doi:10.1002/dvdy.21223

Hakamata, Y., Watanabe, K., Amo, T., Toyomizu, M., and Kikusato, M. (2018). Characterization of mitochondrial content and respiratory capacities of broiler chicken skeletal muscles with different muscle fiber compositions. *J. Poult. Sci.* 55, 210–216. doi:10.2141/jpsa.0170141

Ismail, I., and Joo, S. T. (2017). Poultry meat quality in relation to muscle growth and muscle fiber characteristics. *Korean J. Food Sci. Anim. Resour.* 37, 873–883. doi:10.5851/kosfa.2017.37.6.87

Jackson, H. E., Ono, Y., Wang, X., Elworthy, S., Cunliffe, V. T., and Ingham, P. W. (2015). The role of *SOX6* in zebrafish muscle fiber type specification. *Skelet. Muscle* 5, 2. doi:10.1186/s13395-014-0026-2

Jia, Z., Wang, J., Shi, Q., Liu, S., Wang, W., Tian, Y., et al. (2016). *SOX6* and PDCD4 enhance cardiomyocyte apoptosis through LPS-induced miR-499 inhibition. *Apoptosis.* 21, 174–183. doi:10.1007/s10495-015-1201-6

Jiang, A., Yin, D., Zhang, L., Li, B., Li, R., Zhang, X., et al. (2021). Parsing the microRNA genetics basis regulating skeletal muscle fiber types and meat quality traits in pigs. *Anim. Genet.* 52, 292–303. doi:10.1111/age.13064

Joo, S. T., Kim, G. D., Hwang, Y. H., and Ryu, Y. C. (2013). Control of fresh meat quality through manipulation of muscle fiber characteristics. *Meat Sci.* 95, 828–836. doi:10.1016/j.meatsci.2013.04.044

Lee, C. T., Risom, T., and Strauss, W. M. (2007). Evolutionary conservation of microRNA regulatory circuits: An examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell. Biol.* 26, 209–218. doi:10.1089/dna.2006.0545

Lee, S. H., Joo, S. T., and Ryu, Y. C. (2010). Skeletal muscle fiber type and myofibrillar proteins in relation to meat quality. *Meat Sci.* 86, 166–170. doi:10.1016/j.meatsci.2010.04.040

Li, H., Shu, J., Shan, Y., Chen, W., Song, C., and Xu, W. (2016). Myofiber development during embryonic to neonatal development in duck breeds differing in muscle growth rates. *J. Integr. Agric.* 15, 403–413. doi:10.1016/S2095-3119(14)60949-7

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 30, 923–930. doi:10.1093/bioinformatics/btt656

Lin, S., Lin, X., Zhang, Z., Jiang, M., Rao, Y., Nie, Q., et al. (2018). Copy number variation in *SOX6* contributes to chicken muscle development. *Genes.* 9, 42. doi:10.3390/genes9010042

Liu, J., Liang, X., Zhou, D., Lai, L., Xiao, L., Liu, L., et al. (2016). Coupling of mitochondrial function and skeletal muscle fiber type by a miR-499/*FNIP1*/AMPK circuit. *EMBO Mol. Med.* 8, 1212–1228. doi:10.15252/emmm.201606372

Liu, Y., Sun, Y., Li, Y., Bai, H., Xue, F., Xu, S., et al. (2017). Analyses of Long non-coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Sci. Rep.* 7, 9055. doi:10.1038/s41598-017-08738-9

Liu, Y., Zhang, M., Shan, Y., Ji, G., Ju, X., Tu, Y., et al. (2020). miRNA–mRNA network regulation in the skeletal muscle fiber phenotype of chickens revealed by integrated analysis of miRNAome and transcriptome. *Sci. Rep.* 10, 10619. doi:10.1038/s41598-020-67482-9

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

McCarthy, J. J. (2011). The MyomiR network in skeletal muscle plasticity. *Exerc. Sport Sci. Rev.* 39, 150–154. doi:10.1097/JES.0b013e31821c01e1

Nachtigall, P. G., Dias, M. C., Carvalho, R. F., Martins, C., and Pinhal, D. (2015). MicroRNA-499 expression distinctively correlates to target genes *SOX6* and rod1 profiles to resolve the skeletal muscle phenotype in Nile tilapia. *PLOS ONE.* 10, e0119804. doi:10.1371/journal.pone.0119804

Ohnuki, Y., Umeki, D., Cai, W., Kawai, N., Mototani, Y., Shiozawa, K., et al. (2013). Role of masseter muscle β_2-adrenergic signaling in regulation of muscle activity, myosin heavy chain transition, and hypertrophy. *J. Pharmacol. Sci.* 123, 36–46. doi:10.1254/jphs.12271fp

Pu, F., Xiong, X., Li, Y., Xi, Y., Ma, S., Bai, L., et al. (2022). Transcriptome analysis of oviduct in laying ducks under different stocking densities. *Br. Poult. Sci.* 63, 283–290. doi:10.1080/00071668.2021.1983917

Quiat, D., Voelker, K. A., Pei, J., Grishin, N. V., Grange, R. W., Bassel-Duby, R., et al. (2011). Concerted regulation of myofiber-specific gene expression and muscle performance by the transcriptional repressor *SOX6*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10196–10201. doi:10.1073/pnas.1107413108

Shan, Y., Ji, G., Zhang, M., Liu, Y., Tu, Y., Ju, X., et al. (2022). Use of transcriptome sequencing to explore the effect of *CSRP3* on chicken myoblasts. *J. Integr. Agric.*

Shan, Y., Ji, G., Zou, J., Zhang, M., Tu, Y., Liu, Y., et al. (2020). PGC-1α differentially regulates the mRNA expression profiles of genes related to myofiber type specificity in chicken. *J. Integr. Agric.* 19, 2083–2094. doi:10.1016/S2095-3119(20)63177-X

Shi, Y., Han, Y., Niu, L., Li, J., and Chen, Y. (2019). MiR-499 inhibited hypoxia/reoxygenation induced cardiomyocytes injury by targeting SOX6. *Biotechnol. Lett.* 41, 837–847. doi:10.1007/s10529-019-02685-3

Shu, J., Xiao, Q., Shan, Y., Zhang, M., Tu, Y., Ji, G., et al. (2017). Oxidative and glycolytic skeletal muscles show marked differences in gene expression profile in Chinese Qingyuan partridge chickens. *PLOS ONE.* 12, e0183118. doi:10.1371/journal.pone.0183118

Sibut, V., Hennequet-Antier, C., Le Bihan-Duval, E., Marthey, S., Duclos, M. J., and Berri, C. (2011). Identification of differentially expressed genes in chickens differing in muscle glycogen content and meat quality. *BMC Genomics.* 12, 112. doi:10.1186/1471-2164-12-112

Sirén, J., Välimäki, N., and Mäkinen, V. (2014). Indexing graphs for path queries with applications in genome research. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11, 375–388. doi:10.1109/TCBB.2013.2297101

van Rooij, E., Quiat, D., Johnson, B. A., Sutherland, L. B., Qi, X., Richardson, J. A., et al. (2009). A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev. Cell.* 17, 662–673. doi:10.1016/j.devcel.2009.10.013

Wan, Q., Xu, T., Ding, W., Zhang, X., Ji, X., Yu, T., et al. (2018). miR-499-5p attenuates mitochondrial fission and cell apoptosis via p21 in doxorubicin cardiotoxicity. *Front. Genet.* 9, 734. doi:10.3389/fgene.2018.00734

Wang, X. Y., Chen, X. L., Huang, Z. Q., Chen, D. W., Yu, B., He, J., et al. (2017). MicroRNA-499-5p regulates porcine myofiber specification by controlling *SOX6* expression. *Animal.* 11, 2268–2274. doi:10.1017/S1751731117001008

Xu, M., Chen, X., Chen, D., Yu, B., Li, M., He, J., et al. (2018). MicroRNA-499-5p regulates skeletal myofiber specification via NFATc1/MEF2C pathway and *THRAP1*/MEF2C axis. *Life Sci.* 215, 236–245. doi:10.1016/j.lfs.2018.11.020

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

Zhang, D., Wang, X., Li, Y., Zhao, L., Lu, M., Yao, X., et al. (2014). Thyroid hormone regulates muscle fiber type conversion via miR-133a1. *J. Cell. Biol.* 207, 753–766. doi:10.1083/jcb.201406068

Zhang, Y., Yan, H., Zhou, P., Zhang, Z., Liu, J., and Zhang, H. (2019). MicroRNA-152 promotes slow-twitch myofiber formation via targeting uncoupling protein-3 gene. *Animals.* 9, 669. doi:10.3390/ani9090669

Zhang, Y., Yang, T. L., Li, X., and Guo, Y. (2015). Functional analyses reveal the essential role of *SOX6* and *RUNX2* in the communication of chondrocyte and osteoblast. *Osteoporos. Int.* 26, 553–561. doi:10.1007/s00198-014-2882-3

Zierath, J. R., and Hawley, J. A. (2004). Skeletal muscle fiber type: Influence on contractile and metabolic properties. *PLoS Biol.* 2, e348. doi:10.1371/journal.pbio.0020348

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 October 2022
DOI 10.3389/fgene.2022.1031707

OPEN ACCESS

EDITED BY
Guoqiang Yi,
Chinese Academy of Agricultural Sciences, China

REVIEWED BY
Junhua Liu,
Nanjing Agricultural University, China
Xin Zhou,
Nantong University, China

*CORRESPONDENCE
Jianliang Wu,
wujl@zaas.ac.cn
Yongqing Jiang,
jyq61@sohu.com

†These authors have contributed equally to this work

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

RECEIVED 30 August 2022
ACCEPTED 13 October 2022
PUBLISHED 25 October 2022

CITATION
Zheng K, Guo L, Ullah S, Cao Y, Huang X, shan H, Jiang J, Wu J and Jiang Y (2022), Proteome changes of sheep rumen epithelium during postnatal development. *Front. Genet.* 13:1031707. doi: 10.3389/fgene.2022.1031707

Proteome changes of sheep rumen epithelium during postnatal development

Kaizhi Zheng[1], Liangyong Guo[2], Saif Ullah[3], Yang Cao[1], Xin Huang[1], Huili shan[1], Junfang Jiang[1], Jianliang Wu[1*†] and Yongqing Jiang[1*†]

[1]Institute of Animal Husbandry and Veterinary, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]Huzhou Academy of Agricultural Sciences, Huzhou, China, [3]Faculty of Veterinary and Animal Sciences, Lasbela University of Agriculture Water and Marine Sciences, Lasbela, Pakistan

Background: The development of the rumen epithelium is a critical physiological challenge for sheep. However, the molecular mechanism underlying postnatal rumen development in sheep remains rarely understood.

Results: Here, we used a shotgun approach and bioinformatics analyses to investigate and compare proteomic profiles of sheep rumen epithelium tissue on day 0, 15, 30, 45, and 60 of age. A total of 4,523 proteins were identified, in which we found 852, 342, 164, and 95 differentially expressed proteins (DEPs) between day 0 and day 15, between day 15 and day 30, between day 30 and day 45, between day 45 and day 60, respectively. Furthermore, subcellular localization analysis showed that the DEPs were majorly localized in mitochondrion between day 0 and day 15, after which nucleus proteins were the most DEPs. Finally, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses showed that DEPs significantly enriched in mitochondrion, ubiquitination, histone modifications, glutathione synthase activity, and wnt and nortch signaling pathways.

Conclusion: Our data indicate that the biogenesis of mitochondrion in rumen epithelial cell is essential for the initiation of rumen epithelial development. Glutathione, wnt signaling pathway and nortch signaling pathway participated in rumen epithelial growth. Ubiquitination, post-translational modifications of histone might be key molecular functions in regulating rumen epithelial development.

KEYWORDS

proteome, sheep, rumen, epithelium, postnatal development

1 Introduction

Sheep are a major source of meat, milk and fiber that have a specialized digestive organ, the rumen (Jiang et al., 2014). Rumen is the most important digestive organ for ruminant animal, which accounts for up to 80% of the energetic needs of mature animals (Bergman, 1990). However, the physiological structure and function of fetal rumen are

not well developed during prenatal stage (Arias et al., 1978). For lamb, the transition of milk to solid diets requires the development of a functional rumen. Hence, a fully developed rumen is critical for sheep production.

After birth, the rumen undergoes a series of dramatic ruminal transformations making it to a ruminant, which uses microbial flora to ferment the feed and generates short-chain fatty acid (SCFAs) (Baldwin, 1998). During this transition, the rumen development is consists of three aspects, including morphological development, microbial colonization and functional achievements. For functional achievements, about 85% of SCFAs are directly absorbed by rumen epithelia, facilitating the conversion of low value lignocellulose-rich plant materials to animal protein (Wolin, 1981). While the rumen epithelia also changes to tongue-shaped papillae constituted by stratum basale, stratum spinosum, stratum granulosum and stratum corneum in morphological development, in which stratum spinosum cells played crucial roles in SCFAs metabolism and immune response (Yuan et al., 2022). Followed by the morphological changes and functional achievements of rumen epithelia, great biochemical changes occurs during this stage (Diao et al., 2019), which make rumen basale cell differentiates and migrates to the other types of cells (Steele et al., 2016).

Our previous study has described the dramatic histological changes of lamb rumen papillae from birth to 60 days of age (Zheng et al., 2021). However, we still lack an overview of the changes in protein expression, regulatory factors and protein networks throughout sheep rumen epithelium postnatal development. Recently, proteomics have become a powerful post-genomic tool widely applied to identify and quantify overall proteins present content of a cell, tissue or an organism (Aslam et al., 2017; Keerthikumar, 2017). More and more biologists are using proteomics to reveal single-cell variations, dynamic protein translocations, interaction networks and proteins localizing to multiple compartments (Lundberg and Borner, 2019). Given the central role of rumen epithelium in sheep production, we were promoted to use proteomic as a powerful tool to illuminate the underlying molecular mechanisms that involved in rumen epithelium postnatal development.

2 Material and methods

2.1 Animals

Male Hu Sheep lamb were randomly selected and fed by ewe. From day 10 of age, pellet was used as supplement for lamb. The lamb were sacrificed either at day 0, day 15, day 30, day 45 or day 60 of age with 3 sheep in each age. Slaughtering was carried out through the assembly line of a local commercial abattoir, the slaughtering method followed the traditional procedures, which include stunning. Rumen epithelia were isolated and frozen in liquid nitrogen. Experimental protocols for animal research were approved by the Institutional Animal Care and Use Committees at the Zhejiang Academy of Agricultural Sciences.

2.2 Sample preparation and trypsin digestion for label-free proteome

The sample preparation of protein was performed essentially as the procedure described previously (Liao et al., 2022). Rumen epithelium samples were ground individually in liquid nitrogen and lysed with PASP lysis buffer (100 mM NH4HCO3, 8 M Urea, pH 8.0), followed by 5 min of ultrasonication on ice. The lysate was centrifuged at 12,000 g for 15 min at 4°C. The supernatants were collected and quantified by BCA Protein Assay Kit (Beyotime Institute of Biotechnology, Shanghai, China). 20 μg of the protein sample was loaded to 12% SDS-PAGE gel electrophoresis for quality control and the supernatant was reduced with 10 mM DTT for 1 h at 56°C, and subsequently alkylated with sufficient iodoacetamide (IAM) for 1 h at room temperature in the dark. Then samples were completely mixed with 4 times volume of precooled acetone by vortexing and incubated at −20°C for at least 2 h. Samples were then centrifuged at 12,000 g for 15 min at 4°C and the precipitation was collected.

After washing with 1 ml cold acetone, the pellet was dissolved by dissolution buffer (8 M Urea, 100 mM TEAB, pH 8.5). trypsin and 100 mM TEAB buffer were added, sample was mixed and digested at 37°C for 4 h. Then trypsin and CaCl2 were added digested overnight. Formic acid was mixed with digested sample, adjusted pH under 3, and centrifuged at 12,000 g for 5 min at room temperature. The supernatant was slowly loaded to the C18 desalting column, washed with washing buffer (0.1% formic acid, 3% acetonitrile) 3 times, then added elution buffer (0.1% formic acid, 70% acetonitrile). The eluents of each sample were collected and lyophilized.

2.3 LC-MS/MS analysis

LC-MS/MS analysis was then performed according to the procedure described previously (Zhou et al., 2019). Mobile phase A (0.1% formic acid in H2O) and B solution (0.1% formic acid in 80% acetonitrile) were prepared. The lyophilized powder was dissolved in 10 μl of solution A, centrifuged at 14,000 g for 20 min at 4°C, and 1 μg of the supernatant was injected into a home-made C18 Nano-Trap column (4.5 cm × 75 μm, 3 μm). Peptides were separated in a home-made analytical column (15 cm × 150 μm, 1.9 μm), using a linear gradient elution.

LC-MS/MS analyses were performed by using Q ExactiveTM series mass spectrometer (Thermo Fisher, Germany) at Novogene Co., Ltd. (Beijing, China), with ion source of Nanospray Flex™ (ESI), spray voltage of 2.1 kV and ion transport capillary temperature of 320°C. Full scan range from m/z 350 to 1,500 with resolution of 60,000 (at m/z 200), an automatic gain

control target value was 3 × 106 and a maximum ion injection time was 20 m s. The top 40 precursors of the highest abundant in the full scan were selected and fragmented by higher energy collisional dissociation and analyzed in MS/MS, where resolution was 15,000 (at m/z 200), the automatic gain control target value was 1 × 105, the maximum ion injection time was 45 m s, a normalized collision energy was set as 27%, an intensity threshold was 2.2 × 104, and the dynamic exclusion parameter was 20 s. The raw data of MS detection was named as "raw".

2.4 Protein identification and quantitation

The all resulting spectra were searched by Proteome Discoverer 2.2 (PD 2.2, Thermo). The corresponding proteins were matched with ovis aries database. The search parameters are set as follows: mass tolerance for precursor ion was 10 ppm and mass tolerance for product ion was 0.02 Da. Carbamidomethyl was specified as fixed modifications, Oxidation of methionine was specified as dynamic modification, and acetylation was specified as N-Terminal modification. A maximum of two missed cleavage sites were allowed. To improve the quality of analysis results, the software PD 2.2 further filtered the retrieval results: Peptide Spectrum Matches (PSMs) with a credibility of more than 99% was identified PSMs. The identified protein contains at least 1 unique peptide. The identified PSMs and protein were retained and performed with FDR no more than 1.0%. Principal component analysis (PCA) and volcano plot, which combined fold-change analysis and t-tests, were performed. Proteins with a minimum fold change of 2 (ratio >2 or <0.5, $p < 0.05$) were considered to be regulated differently between comparisons.

2.5 Bioinformatics analysis

GO functional analysis was conducted using the interproscan program against the non-redundant protein database (including Pfam, PRINTS, ProDom, SMART, ProSite, PANTHER), and the databases of Clusters of Orthologous Groups and KEGG (http://www.genome.jp/kegg/) were used to analyze the protein family and pathway. DPEs were used for Volcanic map analysis, cluster heat map analysis and enrichment analysis of GO, subcellular localization and KEGG. The probable protein-protein interactions were predicted using the STRING-db server (http://string.embl.de/).

3 Results

3.1 Protein identification and comparison analysis

To explore the mechanism driving rumen papillae development, we used label-free proteomic strategy to identify different abundant proteins of day 0, day 15, day 30, day 45, and day 60. A total of 4,523 proteins were identified, including 44 proteins uniquely expressed in day 0, 16 proteins uniquely expressed in day 15, 8 proteins uniquely expressed in day 30, 10 proteins uniquely expressed in day 45 and 3 proteins uniquely expressed in day 60 (Figure 1B). To visually differentiate the sample clusters among the observations, PCA analysis was performed to compare the mutual proteins among rumen epithelium tissues of different ages. We found that 66.76% of the variability was explained by the first two principal components, which accounted for 13.88%, and 52.88% of the total variance. The rumen epithelia of day 0, day 15, day 30, and day 60 could be separated completely by identified proteins and distributed in different locations. However, the rumen epithelia of day 45 could not be completely separated from that of day 30 and day 60 (Figure 1A).

The differentially abundant proteins were highlighted by simultaneously considering fold change >2 or <0.5 and $p < 0.05$. A total of 852 proteins were significantly different in comparison of day 0 and day 15 rumen epithelia (Supplementary Table S1). While the number of DEPs was 342 between rumen epithelia of day 15 and day 30 (Supplementary Table S2). Afterwards, this number became 164 between day 30 and day 45 (Supplementary Table S3). And finally, the number DEPs was 95 between rumen epithelia of day 45 and day 60 (Supplementary Table S4). Unsupervised hierarchical clustering of different biological data sets was then conducted as a more rigorous test for the screened proteins to evaluate the rationality, accuracy and dynamic changes of these DEPs (Figure 2).

3.2 Subcellular localization of the DEPs

Between rumen epithelia of day 0 and day 15%, 25.16% of the DEPs were mitochondrion proteins, followed by nucleus proteins (18.01%) and cytoplasm proteins (18.01%) (Figure 3A). Between day 15 and day 30, the majority of DEPs were localized in nucleus (32.27%), while 21.51% were cytoplasm proteins and only 8.76% were mitochondrion proteins (Figure 3B). The DEPs between day 30 and day 45 of age contained 28.81% nucleus proteins, 21.19% cytoplasm proteins and 13.56% mitochondrion proteins (Figure 3C). While Between day 0 and day 15, it contained 27.69% nucleus proteins, 18.46% cytoplasm proteins and 13.85% extracell proteins (Figure 3D).

3.3 Bioinformatics analysis of differentially abundant proteins

In order to assess the major biological processes that participated in rumen epithelial development, we perform a GO annotation to analyze the functional characteristics of

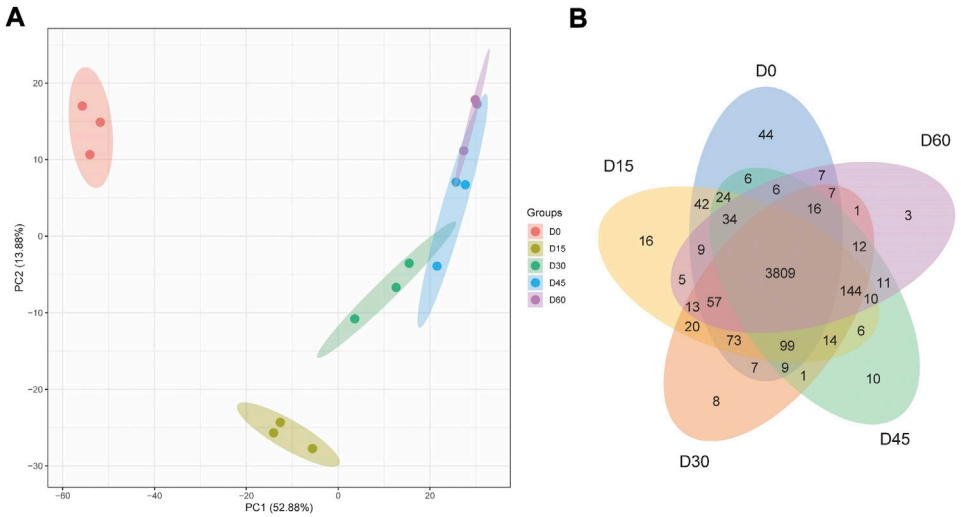


FIGURE 1
PCA scores plot **(A)** and Venn Diagram **(B)** of proteins identified in rumen epithelium at the age of day 0, day 15, day 30, day 45, and day 60.

proteins identified in sheep rumen epithelia of day 0, day 15, day 30, day 45, and day 60 (Figure 4). Bioinformatics analyses were then performed to construct a specific molecular network to explore the biological functions and pathways related to the DEPs from birth to 60 days of age. GO enrichment analysis showed that DEPs in rumen epithelia were significantly enriched in 84 GO terms between day 0 and day 15 ($p < 0.05$). In the biological processes analysis, the top five biological progresses were transmembrane transport (GO:0055085), ATP synthesis coupled proton transport (GO:0015986), ion transport (GO:0006811), monovalent inorganic cation transport (GO:0015672) and hydrogen ion transmembrane transport (GO:1902600). In molecular functions, the top five significant GO terms were hydrogen ion transmembrane transporter activity (GO:0015078), transmembrane transporter activity (GO:0022857), substrate-specific transmembrane transporter activity (GO:0022891), transporter activity (GO:0005215) and ion transmembrane transporter activity (GO:0015075). As for cell components, mitochondrion (GO:0005739), mitochondrial part (GO:0044429), mitochondrial envelope (GO:0005740), organelle envelope (GO:0031967) and intracellular organelle part (GO:0044446) were the top five GO terms that enriched significantly ($p < 0.05$) (Figure 5A, Supplementary Table S5).

While the number of GO enriched terms decreased to 20 between day 15 and day 30 of age. The top five molecular function GO terms were ubiquitin-protein transferase activity (GO:0004842), nucleoside-triphosphate diphosphatase activity (GO:0047429), ferric iron binding (GO:0008199), transferase activity (GO:0016740) and DNA binding (GO:0003677). As for biology processes, cellular iron ion homeostasis (GO:0006879), tRNA modification (GO:0006400), RNA processing (GO:0006396), histone modification (GO:0016570) and macromolecule modification (GO:0043412) were the top five significantly enriched GO terms. And for cell components, the significantly enriched GO terms included DNA-directed RNA polymerase II, holoenzyme (GO:0016591), extracellular matrix (GO:0031012), transferase complex (GO:1990234) and proteinaceous extracellular matrix (GO:0005578) ($p < 0.05$) (Figure 5B, Supplementary Table S6).

Between rumen epithelium tissues of day 30 and day 45, the significantly enriched GO terms of biology progresses were peptidyl-amino acid modification (GO:0018193), peptidyl-lysine modification (GO:0018205), chromatin organization (GO:0006325), chromosome organization (GO:0051276), peptidyl-lysine modification to peptidyl-hypusine (GO:0008612), proton-transporting ATP synthase complex assembly (GO:0043461) and histone acetylation (GO:0016573). As for molecular functions, the GO terms included oxidoreductase activity (GO:0016705), 1-alkyl-2-acetylglycerophosphocholine esterase activity (GO:0003847), IkappaB kinase activity (GO:0008384), histone acetyltransferase activity (GO:0004402) and hydrolase activity (GO:0016788). In cell components, DEPs were enriched in core TFIIH complex (GO:0000439), histone acetyltransferase complex (GO:0000123) and nucleus (GO:0005634) ($p < 0.05$) (Figure 5C, Supplementary Table S7).

Between rumen epithelium tissues of day 45 and day 60, the significantly enriched GO terms of biology progresses were sulfur compound metabolic process (GO:0006790), purine nucleobase biosynthetic process (GO:0009113), glutathione biosynthetic process (GO:0006750) and pointed-end actin

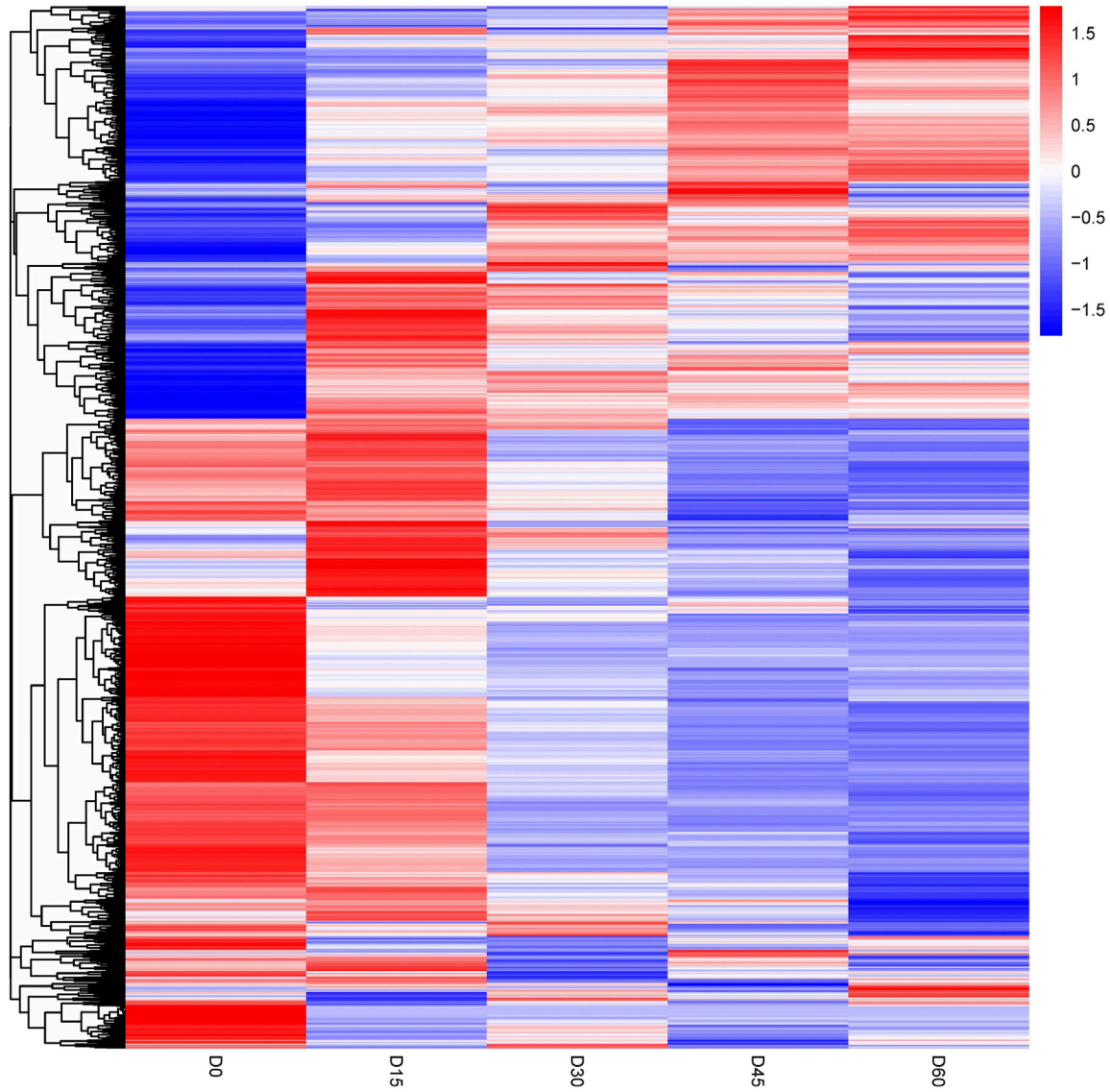


FIGURE 2
Heat map of differentially abundant proteins in rumen epithelium at the age of day 0, day 15, day 30, day 45 and day 60.

filament capping (GO:0051694). As for molecular functions, the GO terms included glutathione synthase activity (GO: 0004363), phosphoribosylamine-glycine ligase activity (GO: 0004637), endopeptidase inhibitor activity (GO:0004866), ligase activity, forming carbon-nitrogen bonds (GO: 0016879) and ubiquitin-protein transferase activity (GO: 0004842), tropomyosin binding (GO:0005523), ubiquitin-ubiquitin ligase activity (GO:0034450) and beta-catenin binding (GO:0008013). In cell components, DEPs were enriched in ubiquitin ligase complex (GO:0000151), spliceosomal complex (GO:0005681) and nuclear part (GO: 0044428) ($p < 0.05$) (Figure 5D, Supplementary Table S8).

3.4 kyoto encyclopedia of genes and genomes pathway enrichment analysis

Then, KEGG pathway enrichment was performed to extract the biological pathways related to the differentially abundant proteins in rumen epithelia from birth to day 60 of age. Between day 0 and day 15, the DEPs were significantly enriched in 33 pathways. The top five significantly enriched KEGG terms were oxidative phosphorylation, Parkinson's disease, metabolic pathways, Huntington's disease and Alzheimer's disease ($p < 0.05$)(Figure 6A, Supplementary Table S9). From day 15 to day 30 of age, the number of significantly

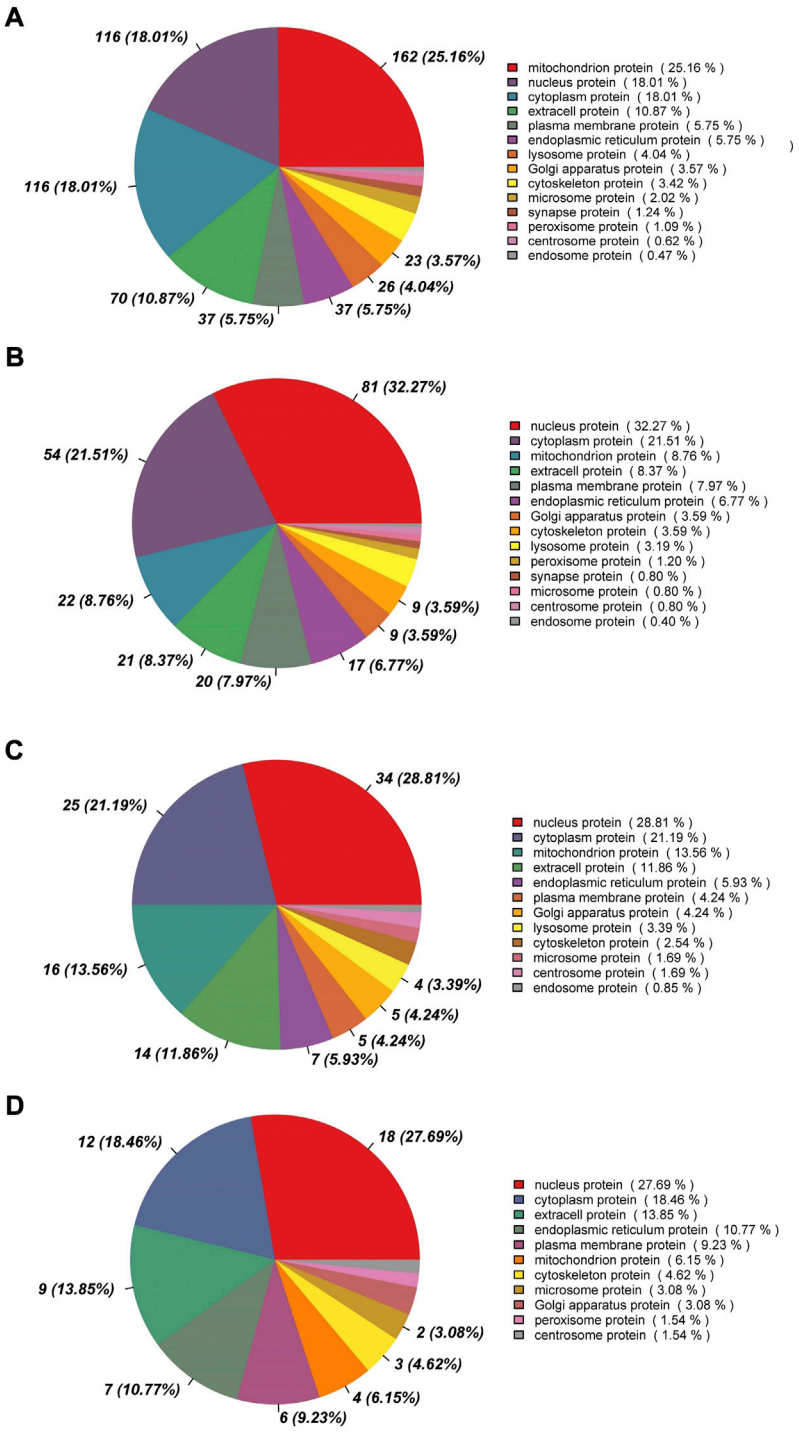


FIGURE 3
Subcellular localization analysis of differentially abundant proteins in rumen epithelium between **(A)** day 0 and day 15, **(B)** day 15 and day 30, **(C)** day 30 and day 45 and **(D)** day 45 and day 60 of age.

enriched KEGG pathways decreased to 9, including cell cycle, other types of O-glycan biosynthesis, cytokine-cytokine receptor interaction, mRNA surveillance pathway, transcriptional misregulation in cancer, ubiquitin mediated proteolysis, Wnt signaling pathway, notch signaling pathway and prostate cancer ($p < 0.05$)(Figure 6B, Supplementary Table S10). From day 30 to

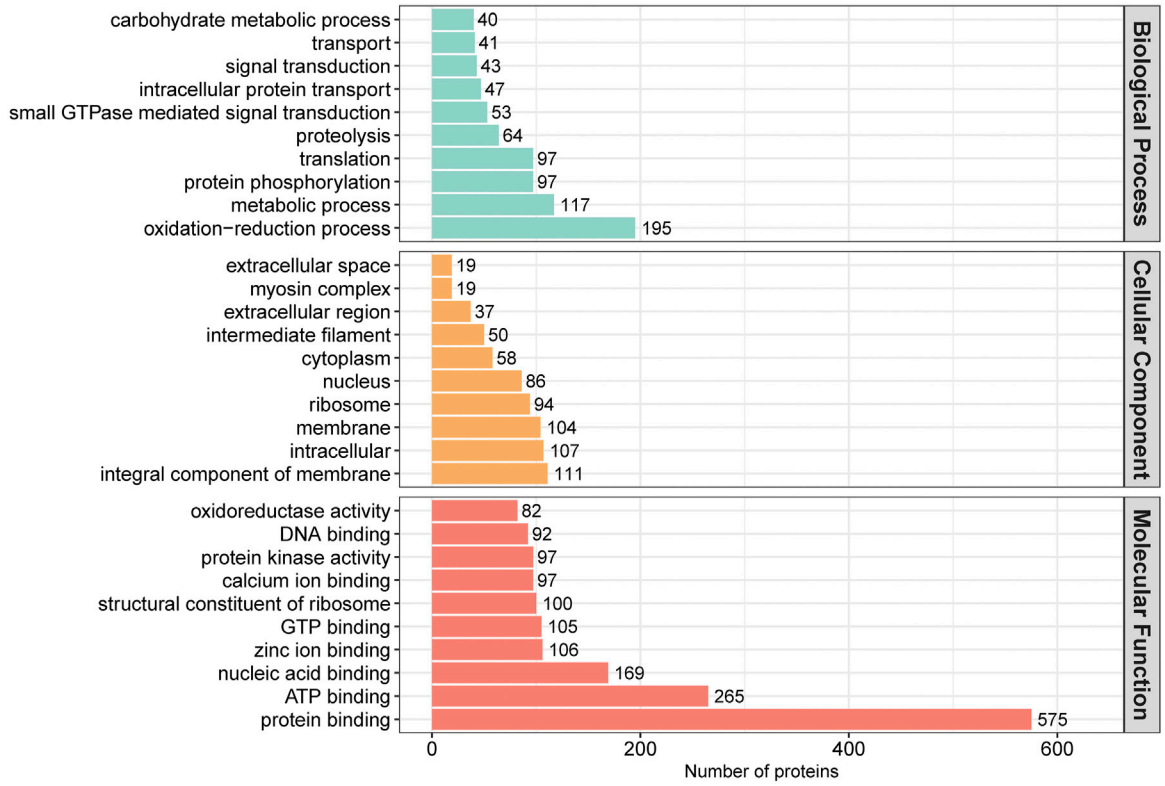


FIGURE 4
GO annotation classification of differentially abundant proteins in rumen epithelium at the age of day 0, day 15, day 30, day 45 and day 60.

day 45 of age, the significantly enriched KEGG pathways were glycine, serine and threonine metabolism, nucleotide excision repair, basal transcription factors, valine, leucine and isoleucine biosynthesis, RNA polymerase and nitrogen metabolism ($p < 0.05$) (Figure 6C, Supplementary Table S11). While, from day 45 to day 60 of age, the significantly enriched KEGG pathways were only cysteine and methionine metabolism (map00270), spliceosome (map03040) and malaria (map05144) ($p < 0.05$)(Figure 6D, Supplementary Table S12).

4 Discussion

The development of the rumen is a critical physiological challenge in newborn ruminants. A better developed rumen epithelium is important for better feed efficiency, which increases profit for producers. In this study, we obtained a comprehensive landscape of proteome profiles across 15 rumen tissue samples during five stages of lamb sucking period, which provided a system overview of proteomic changes of sheep rumen epithelium from birth to weaning.

The high consistence of PCA and heat map analysis results showed that the DEPs upon different ages were distinguishable among rumen epithelium samples of day 0, day 15, day 30, and day 60. While the results of day 45 could not be distinguished from that of day 30 and day 60. Our previous study has found significant histological changes occur on sheep rumen epithelium after birth, which forms tongue shaped papillae at the age of day 15 (Zheng et al., 2021). These data also indicate great biochemical changes occur from birth to day 15 of age in rumen epithelial cells, which might result from dramatic protein expression changes. Consistent with our previous study, 621 proteins were up-regulated and 231 proteins were down-regulated on day 15 of age. After GO and KEGG enrichment analysis, we found that oxidative phosphorylation (OXPHOS) and proton transport were important functions that rumen epithelium gained or enhanced during this period.

The cells of the stratum basale contain mitochondria that contribute to the metabolic properties of the ruminal stratified squamous epithelia (Graham and Simmons, 2005). Currently, researchers have uncovered an important role for mitochondria in the differentiation of the neuronal, hematopoetic, mesenchymal and cardiac systems (Piccoli et al., 2005; Chung et al., 2007; Chen et al., 2008; Khacho et al., 2016). Mitochondrion is a complex organelle that plays essential roles in energy transduction, ATP production and cellular signaling events. In stem cells, differentiation is usually accompanied by a metabolic switch and

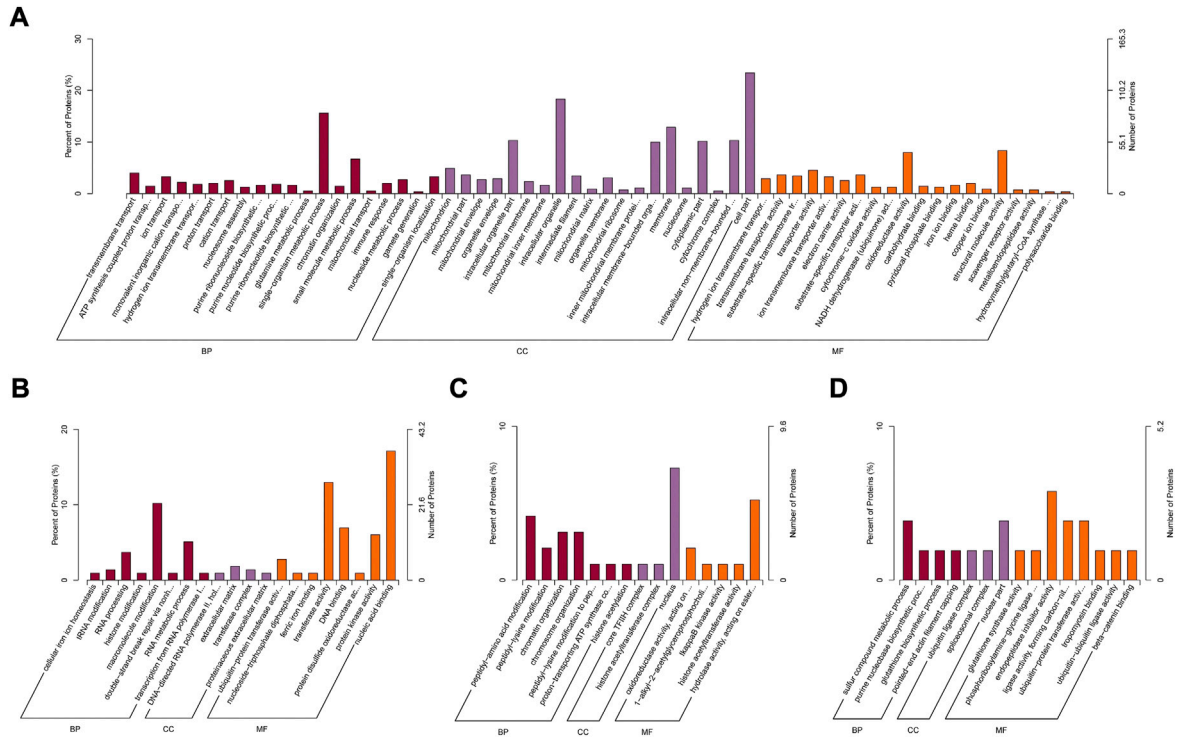


FIGURE 5
GO enrichment analysis of differentially abundant proteins in rumen epithelium between **(A)** day 0 and day 15, **(B)** day 15 and day 30, **(C)** day 30 and day 45 and **(D)** day 45 and day 60 of age. All pathways are listed according to *p*-value.

activation of mitochondrial respiration resulting in the shift from glycolysis to oxidative phosphorylation (Shyh-Chang et al., 2013). Normal differentiated cells primarily rely on mitochondrial OXPHOS to generate the energy for cellular proliferation (Vander Heiden et al., 2009). In rumen epithelium, mitochondria is considered to play a crucial role in the metabolism of short-chain fatty acid by producing ketone bodies, an important circulating source of energy (Leighton et al., 1983). In GO enrichment analysis between day0 and day15 group, we found that mitochondrion and mitochondrial part were top two significantly enriched GO terms, in which more than twenty proteins up-regulated in day 15 of age. While, in KEGG analysis, we also found the up-regulated proteins in rumen epithelia of day 15 significantly enriched in OXPHOS, metabolic pathways, fatty acid metabolism and biosynthesis and citrate cycle when comparing with that of day 0. The subcellular localization analysis also showed that the majority of DEPs were mitochondrion proteins during this stage. These results indicated that the biogenesis of mitochondrion plays a key role in the early development of rumen epithelia. Our results also indicated that the establishment of mitochondrial function in rumen epithelial cell was essential for the initiation of rumen epithelial development. And sheep rumen epithelium might start to gain its function on fatty acid metabolism after the age of day 15.

Proton transport is involved in cellular functions like generation of H+ gradients (Kuijpers and De Pont, 1987). In mammals, generation H+ gradients is of critical importance for mitochondrial energy transduction across mitochondrial inner membrane (Verkhovskaya and Bloch, 2013). Mitochondrion moves protons across its inner membrane during OXPHOS, which converts the energy into high energy phosphates contained within newly generated ATP molecules (Poburko and Demaurex, 2012). Hence, establishment of the proton gradient across the mitochondrial inner membrane is the basis of OXPHOS (Ikon and Ryan, 2017). However, a small proportion of protons directly flow into the mitochondrial matrix across inner mitochondrial membrane without generating ATP, which is called proton leaking (Nanayakkara et al., 2019). Mitochondria are a major source of cellular reactive oxygen species (ROS), which is known as a signaling molecule in various of physiological or pathological conditions (Zhao et al., 2019). The amount of generated ROS in mitochondria determines whether ROS play beneficial or harmful roles. Mitochondrial membrane potential is highly correlated with ROS production, while proton leaking can inhibit the production of ROS by reducing mitochondrial membrane potential (Negre-Salvayre et al., 1997; Suski et al., 2018). Previous studies have reported that proton leaking plays a protective role in early neuronal development (Hoang et al.,

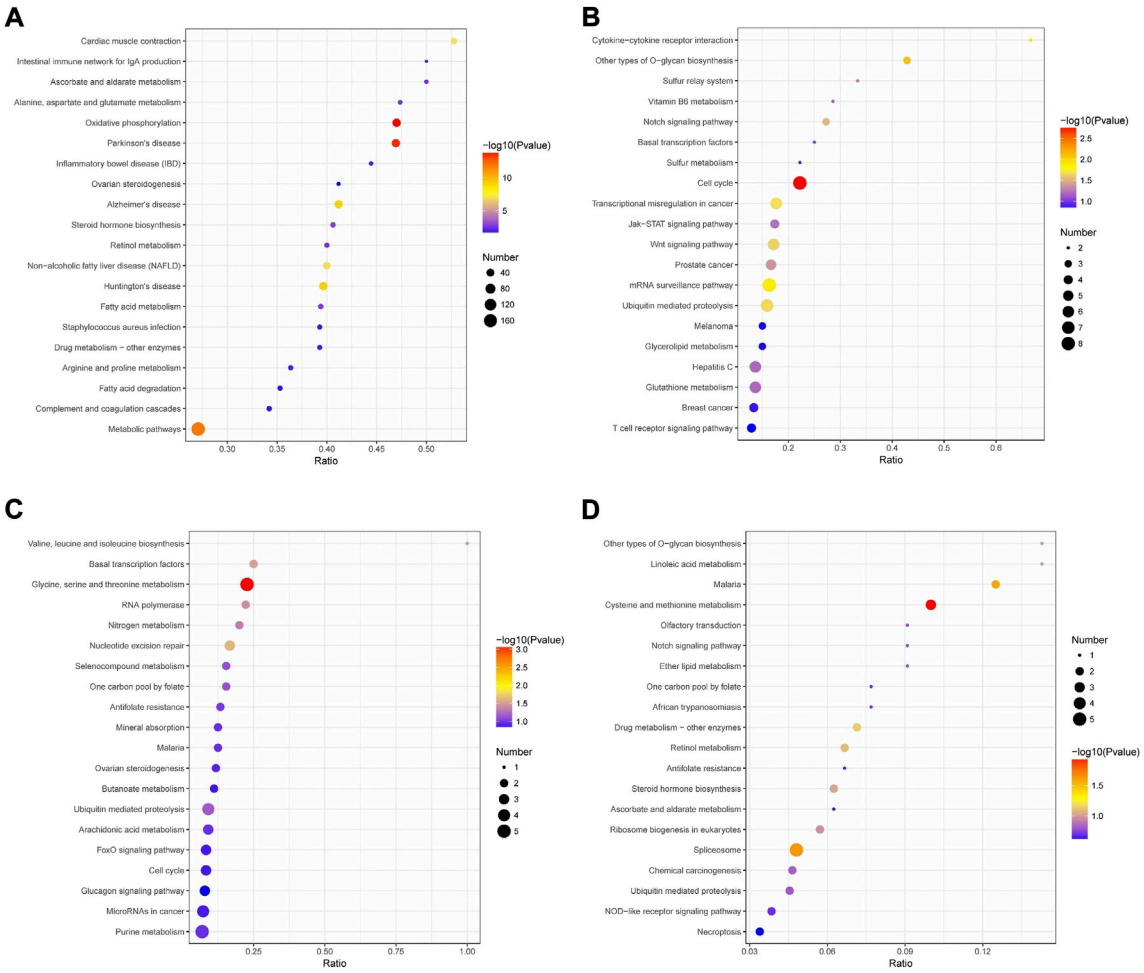


FIGURE 6
KEGG pathway enrichment analysis of differentially abundant proteins in rumen epithelium between **(A)** day 0 and day 15, **(B)** day 15 and day 30, **(C)** day 30 and day 45 and **(D)** day 45 and day 60 of age. All pathways are listed according to *p*-value.

2012). Our results showed that most of the significantly enriched GO terms has either direct or indirect relationship with mitochondrial ion transmembrane transporter activity, especially proton transport, between rumen epithelia of day 0 and day 15. In KEGG analysis between rumen epithelia of day 0 and day 15, we also found DEPs significantly enriched in neurodegenerative disease that related to mitochondria and ROS, like Parkinson's disease, Huntington's disease and Alzheimer's disease (Kim-Han and Dugan, 2005; Jodeiri Farshbaf and Ghaedi, 2017; Tonnies and Trushina, 2017; Shen et al., 2020). The expression of these disease related proteins decreased when proton transport proteins up-regulated on day 15 of age, indicating a protective role of proton transport system in mitochondria of rumen epithelial cell during early rumen development.

In GO enrichment analysis, we found the most significant GO term between rumen epithelia of day 15 and day 30 significantly was ubiquitin-protein transferase activity. Ubiquitination is one of the post-translational modifications that target lysine residue and regulate many cellular processes, including cell division, cell differentiation and immune responses (Popovic et al., 2014). The expression of E3 ubiquitin-protein ligases were down-regulated on day 30. In KEGG enrichment analysis, we also found significant enriched pathways of cell cycle, wnt signaling pathway and notch signaling contained E3 ubiquitin-protein ligases. Wnt signaling is a critical component during embryonic development (Tran and Zheng, 2017), while notch signaling regulates many aspects of metazoan development and tissue renewal (Kopan and Ilagan, 2009). These evidence indicated that rumen epithelial development started since this stage, both wnt signaling and notch signaling were important pathways that participated in this process. And ubiquitination might be a key molecular function in regulating rumen epithelial development.

Post-translational modifications of histone are known to be essential for the regulation of gene transcription, which affect a wide range of biological processes. Histone acetylation is well known as an important regulator of gene expression, which permits switching between permissive and repressive chromatin upon sensing developmental and environmental cues (Shen et al., 2015). Histone acetylation level is controlled by the activity of histone acetyltransferase and deacetylase. The biology function of histone acetyltransferase include cell proliferation, DNA damage repair and glucose metabolism (Poziello et al., 2021). Our previous study has found that the rumen papillae already formed layers of stratum corneum on day 30 of age, while there is no significant histological change except the length of papillae since then (Zheng et al., 2021). In this study, we found DEPs significantly enriched in histone acetylation, and the expression of histone acetyltransferase significantly decreased on day 45 of age. These results indicated that histone acetylation was an important biology process in modulating rumen epithelial development.

The Hu Sheep is naturally weaned around day 60 of age, implying that the function of rumen epithelium has been fully established. In GO enrichment analysis between rumen epithelia of day 45 and day 60, the DEPs were significantly enriched in glutathione synthase activity and glutathione synthetase was up-regulated on day 60 of age. Glutathione is a ubiquitous low molecular weight thiol in eukaryotes, which is required for a diverse range of processes including growth, stress tolerance and cell suicide programs (Diaz-Vivancos et al., 2015). On one hand, it is recruited into the nucleus in early phases of cell proliferation for the synthesis of DNA (Markovic et al., 2007). On the other hand, glutathione is known to remove ROS, which is required for proliferation in response to insulin, fibroblast growth factor, platelet derived growth factor, epidermal growth factor and other growth factors (Burch and Heintz, 2005). These evidence indicated a decrease of rumen epithelial cell proliferation, which is consistent with our previous finding that the ratio of Ki67 positive cell significantly decrease on day 60 of age (Zheng et al., 2021).

To conclude our data indicate that the biogenesis of mitochondrion in rumen epithelial cell is essential for the initiation of rumen epithelial development. Glutathione, Wnt signaling pathway and Notch signaling pathway participated in rumen epithelial growth. And ubiquitination and post-translational modifications of histone might be key molecular function in regulating rumen epithelial development. The proposed datasets provide a useful basis for future studies to better comprehend rumen epithelial development.

Data availability statement

The data presented in the study are deposited in the ProteomeXchange Consortium repository, accession number PXD034787.

Ethics statement

The animal study was reviewed and approved by Institutional Animal Care and Use Committees at the Zhejiang Academy of Agricultural Sciences. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

JW and JJ initiated and directed research; KZ designed and performed experiments, analyzed data and wrote the manuscript; YJ was involved in sample collection, analysis and manuscript edition; HS, SU, YC, and XH were involved in the animal experimentation and sample collection. All authors read and approved the final manuscript.

Funding

This work was supported by Department of Science and Technology of Zhejiang Province, China (Project Nos. LQ20C170002 and 2021C02068-6), Department of Agriculture and Rural Affairs of Zhejiang Province, China (Project No. 2019SNLF016).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1031707/full#supplementary-material

References

Arias, J. L., Cabrera, R., and Valencia, A. (1978). Observations on the histological development of the bovine rumen papillae. Morphological changes due to age. *Anat. Histol. Embryol.* 7 (2), 140–151. doi:10.1111/j.1439-0264.1978.tb00664.x

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., and Rasool, M. H. (2017). Proteomics: Technologies and their applications. *J. Chromatogr. Sci.* 55 (2), 182–196. doi:10.1093/chromsci/bmw167

Baldwin, R. L. t. (1998). Use of isolated ruminal epithelial cells in the study of rumen metabolism. *J. Nutr.* 128, 293S–296S. doi:10.1093/jn/128.2.293S

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70 (2), 567–590. doi:10.1152/physrev.1990.70.2.567

Burch, P. M., and Heintz, N. H. (2005). Redox regulation of cell-cycle re-entry: Cyclin D1 as a primary target for the mitogenic effects of reactive oxygen and nitrogen species. *Antioxid. Redox Signal.* 7 (5-6), 741–751. doi:10.1089/ars.2005.7.741

Chen, C. T., Shih, Y. R., Kuo, T. K., Lee, O. K., and Wei, Y. H. (2008). Coordinated changes of mitochondrial biogenesis and antioxidant enzymes during osteogenic differentiation of human mesenchymal stem cells. *Stem Cells* 26 (4), 960–968. doi:10.1634/stemcells.2007-0509

Chung, S., Dzeja, P. P., Faustino, R. S., Perez-Terzic, C., Behfar, A., and Terzic, A. (2007). Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. *Nat. Clin. Pract. Cardiovasc. Med.* 4, S60–S67. doi:10.1038/ncpcardio0766

Diao, Q., Zhang, R., and Fu, T. (2019). Review of strategies to promote rumen development in calves. *Animals.* 9 (8), E490. doi:10.3390/ani9080490

Diaz-Vivancos, P., de Simone, A., Kiddle, G., and Foyer, C. H. (2015). Glutathione--linking cell proliferation to oxidative stress. *Free Radic. Biol. Med.* 89, 1154–1164. doi:10.1016/j.freeradbiomed.2015.09.023

Graham, C., and Simmons, N. L. (2005). Functional organization of the bovine rumen epithelium. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 288 (1), R173–R181. doi:10.1152/ajpregu.00425.2004

Hoang, T., Smith, M. D., and Jelokhani-Niaraki, M. (2012). Toward understanding the mechanism of ion transport activity of neuronal uncoupling proteins UCP2, UCP4, and UCP5. *Biochemistry* 51 (19), 4004–4014. doi:10.1021/bi3003378

Ikon, N., and Ryan, R. O. (2017). Cardiolipin and mitochondrial cristae organization. *Biochim. Biophys. Acta. Biomembr.* 1859 (6), 1156–1163. doi:10.1016/j.bbamem.2017.03.013

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344 (6188), 1168–1173. doi:10.1126/science.1252806

Jodeiri Farshbaf, M., and Ghaedi, K. (2017). Huntington's disease and mitochondria. *Neurotox. Res.* 32 (3), 518–529. doi:10.1007/s12640-017-9766-1

Keerthikumar, S. (2017). An introduction to proteome bioinformatics. *Methods Mol. Biol.* 1549, 1–3. doi:10.1007/978-1-4939-6740-7_1

Khacho, M., Clark, A., Svoboda, D. S., Azzi, J., MacLaurin, J. G., Meghaizel, C., et al. (2016). Mitochondrial dynamics impacts stem cell identity and fate decisions by regulating a nuclear transcriptional program. *Cell Stem Cell* 19 (2), 232–247. doi:10.1016/j.stem.2016.04.015

Kim-Han, J. S., and Dugan, L. L. (2005). Mitochondrial uncoupling proteins in the central nervous system. *Antioxid. Redox Signal.* 7 (9-10), 1173–1181. doi:10.1089/ars.2005.7.1173

Kopan, R., and Ilagan, M. X. (2009). The canonical notch signaling pathway: Unfolding the activation mechanism. *Cell* 137 (2), 216–233. doi:10.1016/j.cell.2009.03.045

Kuijpers, G. A., and De Pont, J. J. (1987). Role of proton and bicarbonate transport in pancreatic cell function. *Annu. Rev. Physiol.* 49, 87–103. doi:10.1146/annurev.ph.49.030187.000511

Leighton, B., Nicholas, A. R., and Pogson, C. I. (1983). The pathway of ketogenesis in rumen epithelium of the sheep. *Biochem. J.* 216 (3), 769–772. doi:10.1042/bj2160769

Liao, R., Xia, Q., Zhou, C., Geng, F., Wang, Y., Sun, Y., et al. (2022). LC-MS/MS-based metabolomics and sensory evaluation characterize metabolites and texture of normal and spoiled dry-cured hams. *Food Chem.* 371, 131156. doi:10.1016/j.foodchem.2021.131156

Lundberg, E., and Borner, G. H. H. (2019). Spatial proteomics: A powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* 20 (5), 285–302. doi:10.1038/s41580-018-0094-y

Markovic, J., Borras, C., Ortega, A., Sastre, J., Vina, J., and Pallardo, F. V. (2007). Glutathione is recruited into the nucleus in early phases of cell proliferation. *J. Biol. Chem.* 282 (28), 20416–20424. doi:10.1074/jbc.M609582200

Nanayakkara, G. K., Wang, H., and Yang, X. (2019). Proton leak regulates mitochondrial reactive oxygen species generation in endothelial cell activation and inflammation - a novel concept. *Arch. Biochem. Biophys.* 662, 68–74. doi:10.1016/j.abb.2018.12.002

Negre-Salvayre, A., Hirtz, C., Carrera, G., Cazenave, R., Troly, M., Salvayre, R., et al. (1997). A role for uncoupling protein-2 as a regulator of mitochondrial hydrogen peroxide generation. *FASEB J.* 11 (10), 809–815. doi:10.1096/fasebj.11.10.9271366

Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50 (D1), D543–D552. doi:10.1093/nar/gkab1038

Piccoli, C., Ria, R., Scrima, R., Cela, O., D'Aprile, A., Boffoli, D., et al. (2005). Characterization of mitochondrial and extra-mitochondrial oxygen consuming reactions in human hematopoietic stem cells. Novel evidence of the occurrence of NAD(P)H oxidase activity. *J. Biol. Chem.* 280 (28), 26467–26476. doi:10.1074/jbc.M500047200

Poburko, D., and Demaurex, N. (2012). Regulation of the mitochondrial proton gradient by cytosolic Ca^{2+} signals *Pflugers Arch.* 464 (1), 19–26. doi:10.1007/s00424-012-1106-y

Popovic, D., Vucic, D., and Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. *Nat. Med.* 20 (11), 1242–1253. doi:10.1038/nm.3739

Poziello, A., Nebbioso, A., Stunnenberg, H. G., Martens, J. H. A., Carafa, V., and Altucci, L. (2021). Recent insights into histone acetyltransferase-1: Biological function and involvement in pathogenesis. *Epigenetics* 16 (8), 838–850. doi:10.1080/15592294.2020.1827723

Shen, Y., Wei, W., and Zhou, D. X. (2015). Histone acetylation enzymes coordinate metabolism and gene expression. *Trends Plant Sci.* 20 (10), 614–621. doi:10.1016/j.tplants.2015.07.005

Shen, Y., Wu, Q., Shi, J., and Zhou, S. (2020). Regulation of SIRT3 on mitochondrial functions and oxidative stress in Parkinson's disease. *Biomed. Pharmacother.* 132, 110928. doi:10.1016/j.biopha.2020.110928

Shyh-Chang, N., Daley, G. Q., and Cantley, L. C. (2013). Stem cell metabolism in tissue development and aging. *Development* 140 (12), 2535–2547. doi:10.1242/dev.091777

Steele, M. A., Penner, G. B., Chaucheyras-Durand, F., and Guan, L. L. (2016). Development and physiology of the rumen and the lower gut: Targets for improving gut health. *J. Dairy Sci.* 99 (6), 4955–4966. doi:10.3168/jds.2015-10351

Suski, J., Lebiedzinska, M., Bonora, M., Pinton, P., Duszynski, J., and Wieckowski, M. R. (2018). Relation between mitochondrial membrane potential and ROS formation. *Methods Mol. Biol.* 1782, 357–381. doi:10.1007/978-1-4939-7831-1_22

Tonnies, E., and Trushina, E. (2017). Oxidative stress, synaptic dysfunction, and Alzheimer's disease. *J. Alzheimers Dis.* 57 (4), 1105–1121. doi:10.3233/JAD-161088

Tran, F. H., and Zheng, J. J. (2017). Modulating the wnt signaling pathway with small molecules. *Protein Sci.* 26 (4), 650–661. doi:10.1002/pro.3122

Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009). Understanding the warburg effect: The metabolic requirements of cell proliferation. *Science* 324 (5930), 1029–1033. doi:10.1126/science.1160809

Verkhovskaya, M., and Bloch, D. A. (2013). Energy-converting respiratory complex I: On the way to the molecular mechanism of the proton pump. *Int. J. Biochem. Cell Biol.* 45 (2), 491–511. doi:10.1016/j.biocel.2012.08.024

Wolin, M. J. (1981). Fermentation in the rumen and human large intestine. *Science* 213 (4515), 1463–1468. doi:10.1126/science.7280665

Yuan, Y., Sun, D. M., Qin, T., Mao, S. Y., Zhu, W. Y., Yin, Y. Y., et al. (2022). Single-cell transcriptomic landscape of the sheep rumen provides insights into physiological programming development and adaptation of digestive strategies. *Zool. Res.* 43 (4), 634–647. doi:10.24272/j.issn.2095-8137.2022.086

Zhao, R. Z., Jiang, S., Zhang, L., and Yu, Z. B. (2019). Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *Int. J. Mol. Med.* 44 (1), 3–15. doi:10.3892/ijmm.2019.4188

Zheng, K., Wu, J., Guo, L., Ying, Y., Li, P., Cao, Y., et al. (2021). The involvement of translationally controlled tumor protein during lamb rumen epithelium development. *Acta Histochem.* 123 (5), 151737. doi:10.1016/j.acthis.2021.151737

Zhou, C. Y., Wang, C., Tang, C. B., Dai, C., Bai, Y., Yu, X. B., et al. (2019). Label-free proteomics reveals the mechanism of bitterness and adhesiveness in Jinhua ham. *Food Chem.* 297, 125012. doi:10.1016/j.foodchem.2019.125012

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 November 2022
DOI 10.3389/fgene.2022.1021103

OPEN ACCESS

EDITED BY
Guoqiang Yi,
Agricultural Genomics Institute at Shenzhen (CAAS), China

REVIEWED BY
Zuo Bo,
Huazhong Agricultural University, China
Yongzhen Huang,
Northwest A&F University, China

*CORRESPONDENCE
Yang Cao,
caoyang003@163.com

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

RECEIVED 17 August 2022
ACCEPTED 16 September 2022
PUBLISHED 22 November 2022

CITATION
Cao Y, Yu Y, Zhang L, Liu Y, Zheng K, Wang S, Jin H, Liu L and Cao Y (2022), Transcript variants of long-chain acyl-CoA synthase 1 have distinct roles in sheep lipid metabolism.
Front. Genet. 13:1021103.
doi: 10.3389/fgene.2022.1021103

Transcript variants of long-chain acyl-CoA synthase 1 have distinct roles in sheep lipid metabolism

Yang Cao[1,2], Yongsheng Yu[1], Lichun Zhang[1], Yu Liu[1], Kaizhi Zheng[2], Sutian Wang[3], Haiguo Jin[1], Lixiang Liu[1] and Yang Cao[1]*

[1]Institute of Animal Biotechnology, Jilin Academy of Agricultural Science, Gongzhuling, China, [2]Institute of Animal Husbandry and Veterinary, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [3]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, China

Mutton has recently been identified to be a consumer favorite, and intermuscular fat is the key factor in determining meat tenderness. Long-chain acyl-CoA synthetase 1 (ACSL1) is a vital subtype of the ACSL family that is involved in the synthesis of lipids from acyl-CoA and the oxidation of fatty acids. The amplification of the ACSL1 gene using rapid amplification of cDNA ends revealed that the alternative polyadenylation (APA) results in two transcripts of the ACSL1 gene. Exon 18 had premature termination, resulting in a shorter CDS region. In this study, the existence of two transcripts of varying lengths translated normally and designated ACSL1-a and ACSL1-b was confirmed. Overexpression of ACSL1-a can promote the synthesis of an intracellular diglyceride, while ACSL1-b can promote triglyceride synthesis. The transfection of ACSL1 shRNA knocks down both the transcripts, the triglyceride content was significantly reduced after differentiation and induction; and lipidome sequencing results exhibited a significant decrease in 14–22 carbon triglyceride metabolites. The results of the present study indicated that the ACSL1 gene played a crucial role in the synthesis of triglycerides. Furthermore, the two transcripts involved in various interactions in the triglyceride synthesis process may be the topic of interest for future research and provide a more theoretical basis for sheep breeding.

KEYWORDS

ACSL1, transcript variants, sheep, APA, lipid metabolism

1 Introduction

Sheep are widely reared as herbivorous livestock all over the world. Mutton is a high protein and low cholesterol meat, which contains more Ω-3 polyunsaturated fatty acids, monounsaturated fatty acids, single-chain fatty acids, and conjugated linoleic acid (Chikwanha et al., 2017). These fatty acids are associated with human health and can provide essential fatty acids in a healthy human diet, which has been welcomed by consumers in recent years. Intermuscular fat is the key factor in determining meat tenderness and is the primary component of marbling in meat products that consumers

appreciate (Crouse et al., 1984). However, excess fat accumulation will result in feed waste and animal burden. Therefore, a better understanding of adipose tissue metabolic rules can improve production efficiency.

Long-chain acyl-CoA synthetases (ACSLs) are crucial for the synthesis of acyl-CoA while ACSLs can activate fatty acids guiding them into specific metabolic pathways (Ellis et al., 2010). The catalyzed acyl-CoA was involved in different pathways; however, there is no evidence to prove that acyl-CoA is linked to a specific pathway. ACSL1 and ACSL5 both are involved in the production of acyl-CoA, which is primarily used in triglyceride synthesis and mitochondrial oxidation. The ACSLs are saturated and unsaturated fatty acids with 8–22 carbon triglycerides, in addition; each subtype has a different substrate preference. The first one, ACSL1, prefers saturated and monounsaturated fatty acids with 16–18 carbon triglycerides (Suzuki et al., 1990). The subtypes of ACSLs exist with specific abundances in different tissues. ACSL1 is more abundant in the liver and adipose tissues (Paul et al., 2014), and the ester acyl-CoA is essential for triglyceride synthesis (Pan et al., 2010). ACSL1 overexpression was found to promote oleic acid infiltration into triglycerides in NIH-3T3 fibroblasts and PC12 neurons (Souza et al., 2002; Marszalek et al., 2004). Post-overexpression of ACSL1 in mice hearts resulted in a 12-fold increase in triglyceride content and a 1.5-fold increase in choline glycerin phospholipid content (Chiu et al., 2001). During 3T3-L1 adipocyte differentiation, a four-fold increase was observed in the ACSL1 expression, in addition to this, no change was found in the expression levels of other subtypes indicating ACSL1 as the major player in promoting 3T3-L1 adipocytes' triglyceride synthesis (Yue, 2013). Previous studies mentioned that the increase in the activity of acyl-CoA can be achieved by overexpressing ACSL1 in primary hepatocytes of rats. However, ACSL1 increased oleic acid infiltration into phospholipids and diglycerides, but did not promote oleic acid infiltration into triglycerides (Gluchowski et al., 2017). ACSL1 contains multiple promoters, consequently resulting in multiple transcripts with distinct functions in different tissues (Suzuki et al., 1995). The different transcripts of the ACSL1 gene may be responsible for the various roles of the ACSL1 gene in triglyceride synthesis.

Alternative polyadenylation (APA) is important in regulating gene stability, translation, and transport (Sandberg et al., 2008; Mayr and Bartel, 2009). In addition to playing a regulatory role in the gene expression process, APA can result in transcripts with varying lengths of 3′UTR. APA is a ubiquitous phenomenon in transcription that has been discovered in more than 50% of human and 30% of mice genes (Tian et al., 2005). APA regulates gene expression, thereby regulating specific biological functions. APA is classified into two types among which the first one includes the normal termination of transcription. Polyadenylation occurs at the 3′end. Multiple polyA sites are located downstream of the stop codon, resulting in the formation of transcripts with different lengths encoding the same protein, this type of 3′UTR is considered a tandem 3′UTR. The second one occurs at different poly-A sites located in different exons or introns, resulting in coding regions of varying lengths that encode different proteins (Sun et al., 2012).

Previous experiments compared the production performance of Duhan hybrid sheep (Dorper sheep × Small-tailed Han sheep F1) and Small-tailed Han sheep and observed the significantly higher average daily gain of hybrid sheep than that of Small-tailed Han sheep, whereas the comparatively increased intramuscular fat and oleic acid were also reported for hybrid sheep. Using MeDIP-Seq (methylated DNA immunoprecipitation sequencing), two populations of Duhan sheep and Small-tailed Han sheep were sequenced for whole-genome methylation. The results demonstrated a significant difference between expression levels of ACSL1 when two different groups were analyzed ($p < 0.05$), and the ACSL1 gene is regulated by methylation, which affects lipid metabolism and meat quality (Cao et al., 2017). Although the sheep ACSL1 gene sequence is unclear, the existence of multiple transcript variants within this gene is an intriguing area to investigate further.

2 Materials and methods

2.1 Materials

The tissues used for the experiment were isolated from 40-day-old Small-tailed Han sheep raised by the Jilin Academy of Agricultural Sciences. Also, the cells were stored in the Jilin Academy of Agricultural Sciences.

2.2 Rapid amplification of cDNA ends

Primers were designed according to the principles of rapid amplification of cDNA ends (RACE) primer design using ACSL1 gene sequences from GenBank. Primer Premier 6.0 software was used to design 3′UTR primers and labeled as SP1, SP2, SP3, and SP4 (Table 1).

Reverse transcription, 5′-end phosphorylation of primers, and ligation assays were performed according to modified protocols of the manufacturers of SMARTer RACE Kit (Clontech). Tissue RNA was extracted and reverse transcribed with specific primers followed by nested PCR with the primers listed in Table 1.

The target fragment was amplified according to the RACE system. PCR reaction cycles are as follows: 94°C for 30 s, 72°C for 3 min, five cycles; 94°C for 30 s, 70°C for 30 s, 72°C for 3 min, five cycles; 94°C for 30 s, 68°C for 30 s, 72°C for 3 min, 25 cycles. After ligating with pmd18-t vector, the fragment was sequenced by a biological company.

TABLE 1 Sequence of primers.

Primer name	Sequence/(5′-3′)	Product size/bp
SP1	CCTTGGCAGCCAGATAATTCA	450 bp
SP2	GGACAAGCAAACACCACGCTGA	
SP3	CTGGCACAAGGGGAGTACATAGCTCC	750 bp
SP4	TTCGGAATTATTTCAGGTCACAGATCGATG	1,570 bp
ACSL1-a-F	ATGATGCAAGCCCACGAGCTGT	2,100 bp
ACSL1-a-R	TTAGACTTTGATGGTGGAGTAA	
ACSL1-b-F	ATGATGCAAGCCCACGAG	1,836 bp
ACSL1-b-R	TCAGGCACCTAGTGTCT	
NB-F	CCCTTGTTTAGGCTCTCGG	471 bp
NB-R	TTGGGCTTCTGTCTGTTGG	
ACSL1-YF1	GCCATCACCTACATCATCAACAA	151 bp
ACSL1-YR1	ACACTTCTTGCCTCGTTCCA	
ACSL1-YF2	CAGAAACAAGGATGTCAAAAAA	120 bp
ACSL1-YR2	GAGTTCAGGGTGGAGATAGATG	
β-actin-F	GTCCACCTTCCAGCAGAT	96 bp
β-actin-R	GCTAACAGTCCGCCTAGAA	
PPARγ-F	GAGCTGACCCGATGGTT	150 bp
PPARγ-R	TGAGGGAGTTGGAAGGC	
CEBPα-F	CGTGGAGACGCAACAGAAG	105 bp
CEBPα-R	AAGATGCCCCGCAGTGT	
LPL-F	GCCAAAAGAAGCAGCAGCAAG	178 bp
LPL-R	GCAGGGTAAAAGGGATGTT	

Based on the RACE result, primers F1, R1, F2, and R2 (Table 1) were used to clone the CDS regions of two different transcripts of the sheep ACSL1 gene. ACSL1-a was amplified with primers F1 and R1, while amplification of ACSL1-b was carried out with F2 and R2. A biological company sequenced the target fragment that had been ligated with the pmd18-t vector.

2.3 Northern blotting

Comparing the sequences of the two transcripts led to the selection of the common sequence as the target. Probe primers (NB-F and NB-R in Table 1) were designed using the software Primer Premier 6.0.

The probes were prepared using PCR with a DIG marker. Isolation of total RNA from the heart, liver muscle, and adipocytes was carried out *via* TRIzol. After 1% formaldehyde denaturing gel electrophoresis, the membrane was transferred using the upward capillary method for 20 h and then fixed at 80°C for 2 h. The denatured probe was used for overnight hybridization at 50°C after a 2 h prehybridization. The membrane was rinsed at room temperature after hybridization, and the machine was exposed for 3 h to detect the signal.

2.4 Western blotting

The cells were digested with trypsin and then collected after a short centrifugation. Lysis buffer (RIPA, Thermo Scientific) was added to the lysate cells and protein concentration was determined using an Enhanced BCA Protein Assay Kit (Beyotime). In a proportionally diluted sample, 5 × loading buffer was added. Denaturation was carried out at 98°C for 10 min using 10% SDS-PAGE gel electrophoresis, and then 80 V electrophoresis for 30 min was performed which was switched to 120 V electrophoresis for 2 h. In an ice box, the film was transferred for 90 min at 200 mA. The polyvinylidene difluoride (PVDF) membranes were washed thrice with TBST (Tris-buffered saline with 0.1% Tween 20 detergent) and then blocked with 5% skimmed milk at room temperature for 2 h. The membranes were then incubated with a primary antibody. The corresponding species were incubated with the secondary antibody solution at room temperature for 2 h. The film was then rinsed with TBST thrice and the strip was detected using an ECL-PLUS kit (Beyotime).

2.5 Real-time PCR

Primers were designed to detect transcript expression in two sheep tissue samples based on the CDS region derived

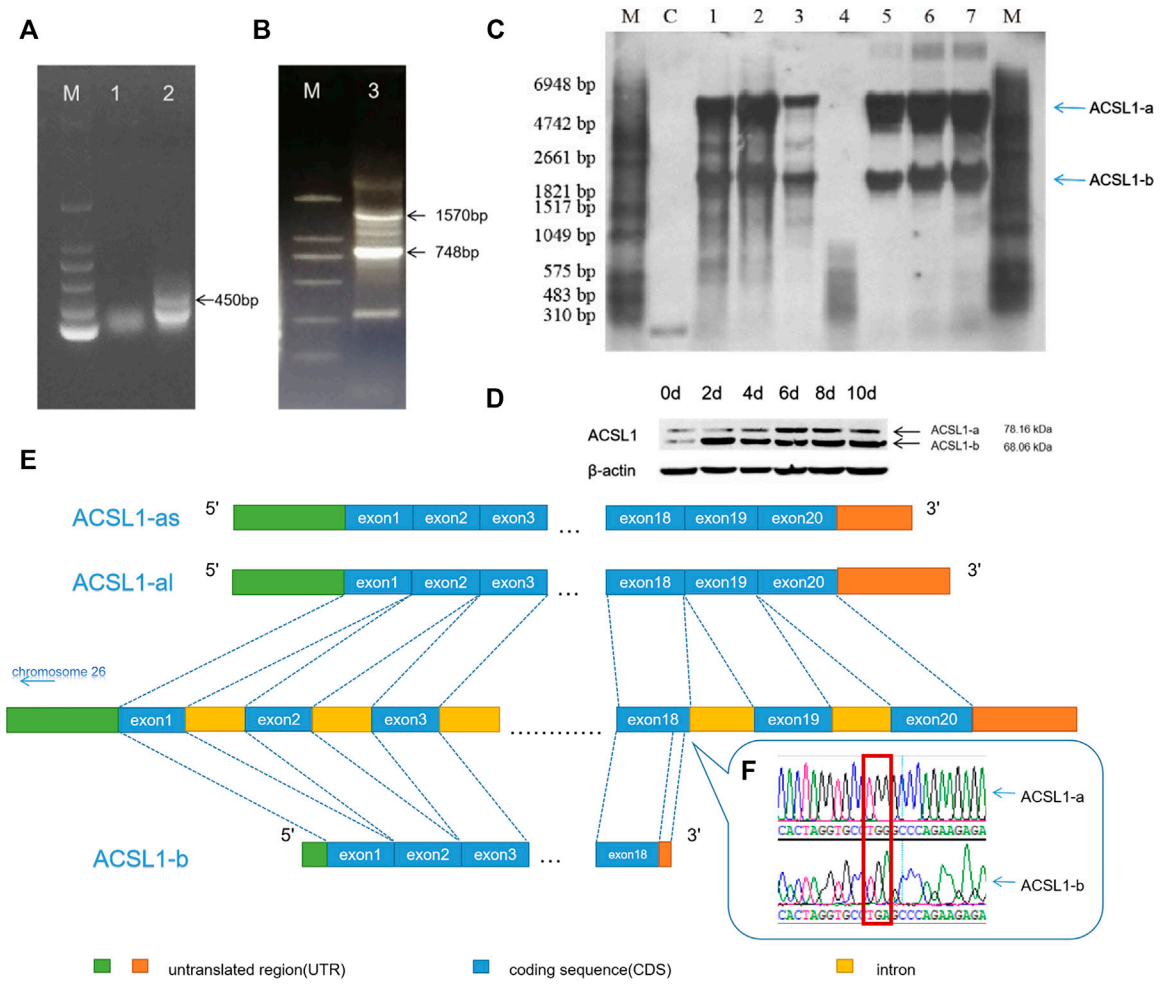


FIGURE 1
(A) 3′RACE PCR result of the sheep ACSL1 gene of SP1 and SP2. **(B)** 3′RACE PCR result of the sheep ACSL1 gene of SP3 and SP4 (lane M: DL2000 marker; lane 1: 3′RACE by SP1; lane 2: 3′RACE by SP2; lane 3: 3′RACE by SP3 and SP4). **(C)** Northern blotting results of sheep tissues and adipocytes of different differentiation stages (lane M: RNA molecular weight marker; lane C: positive control; lane 1: heart; lane 2: liver; lane 3: muscle; lane 5: sheep preadipocytes on the day 1 of differentiation; lane 6: sheep preadipocytes on the day 4 of differentiation; lane 7: the induced differentiation of sheep preadipocytes on day 8). **(D)** Western blotting results of sheep tissues and adipocytes of different differentiation stages (when cells were added to the induction medium, it was taken as day 0). **(E)** Schematic diagram of different transcripts of the ACSL1 gene. **(F)** RACE sequencing results of G–A mutation.

from a cloned ACSL1 gene. The other primers were designed based on sequences on the NCBI (PPARγ ID: XM_012099243.2, C/EBPα ID:NM_001308574.1, LPL ID: NM_001009394.1).

Total RNA from the heart, liver, spleen, lung, kidney, intestine, adipose, and muscle was extracted using the TRIzol method and reverse transcribed into cDNA (Takara). Quantitative real-time PCR (qRT-PCR) amplification was performed with SYBR Premix Ex Taq using Roche 480 LightCycler. The target gene expression was analyzed using the $2^{-\Delta\Delta Ct}$ method and normalized to β-actin.

2.6 Immunofluorescence

Preadipocytes were sown in a 24-well plate (10^4 cells per well) for cell culturing and counting. The next day, a 10-fold increase in mitochondrial dye was applied to preadipocytes. Fluorescence was detected using a fluorescence microscope after 16 h. After fixing the cells at room temperature for 20 min, the fixative solution was rinsed and sealed with the blocking solution for 15 min. The primary antibody (ACSL1, Bioss) was diluted in the proportion of 1:500 and further incubated at 4°C overnight. The cells were kept at room

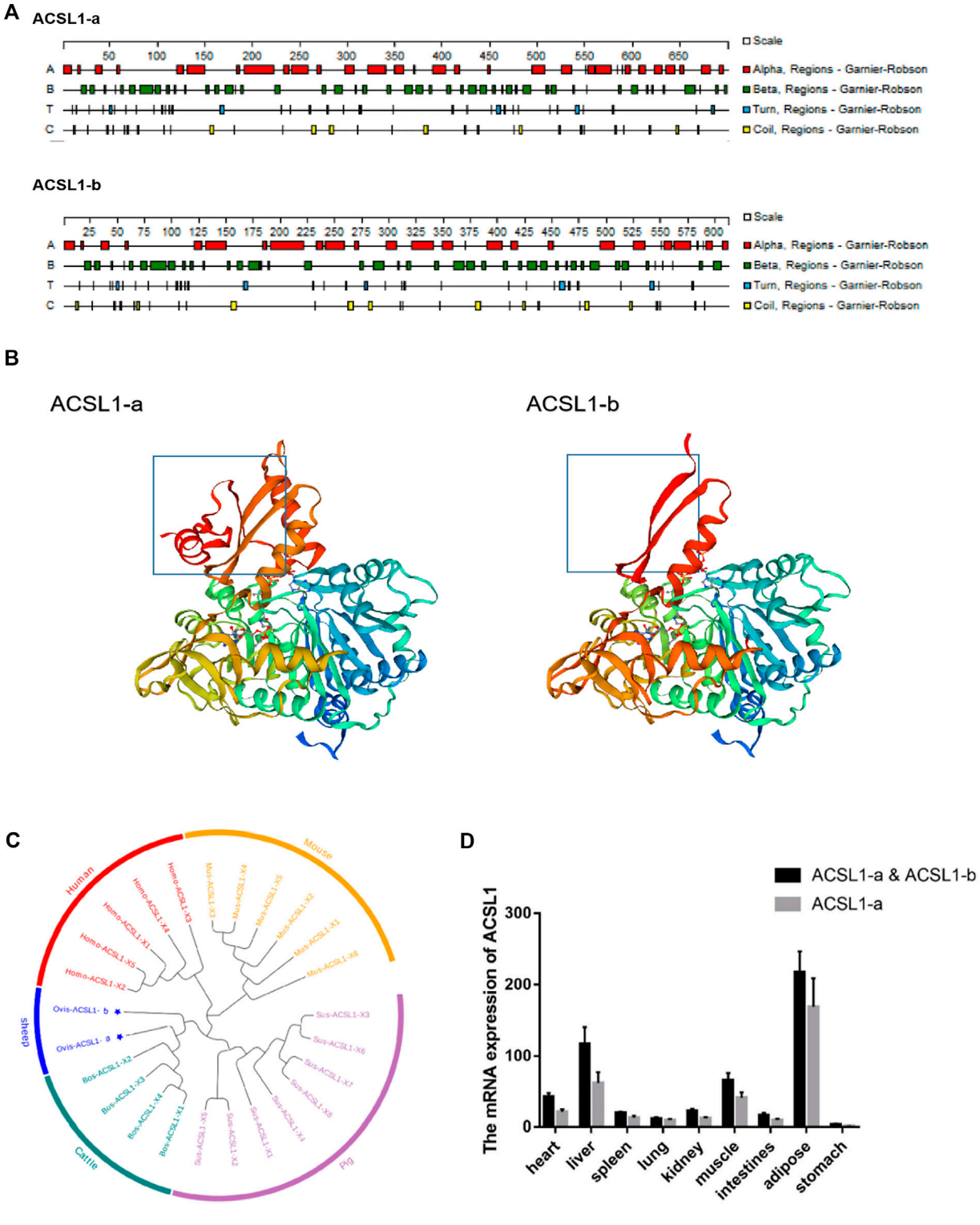


FIGURE 2
(A) Secondary structures of ACSL1-a and ACSL1-b proteins. **(B)** Tertiary structures of ACSL1-a and ACSL1-b proteins. **(C)** Phylogenetic tree depicting different ACSL1 transcripts. **(D)** Expression result of ACSL1 gene expression in various tissues.

TABLE 2 Prediction of sheep ACSL1 subcellular localization.

Location	ACSL1-a (%)	ACSL1-b
Mitochondrial	52.2	69.6%
Nuclear	43.5	30.4%
Cytoplasmic	4.3	—

temperature for 1 h the next day before being rinsed three times with phosphate-buffered saline (PBS). The secondary antibody (anti-rabbit, CST) was diluted in 1: 1,000 proportions and further incubated at room temperature for 2 h. The cells were then washed thrice with PBS. DAPI (4,6-diamidino-2-phenylindole) was diluted in proportion and stained at room temperature for 15 min. Superfluous dye was rinsed away before examining the fluorescence under the microscope.

2.7 Bioinformatics analysis

ORF finder (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi) was used to determine the coding region of the gene, UCSC (http://genome.ucsc.edu/) was utilized to analyze the exon, prediction of the subcellular location of the protein was conducted using PSORT II (https://www.genscript.com/tools/psort). Protean (DNASTAR) was used to predict the secondary structure of the ACSL1 protein, while the tertiary structure of the ACSL1 protein was predicted *via* SWISS-MODEL (http://swissmodel.expasy.org/).

2.8 Cell transfection and sample collection

The shACSL1-297 (GGGCATACAGGTGTCCAATAA) was synthesized and ligated to the pGPU6 vector for sequencing. The recombinant shRNA vector (1 μg) of ACSL1 was transfected into 1 × 10^6 preadipocytes using FuGENE (omega). On day 8, fully differentiated cells that had matured into adipocytes were collected. In the lipid metabolome analysis, at least 10^7 cells were used. The trypsinized cells were then transferred into an EP tube. The supernatant was removed, and the cells were rinsed with pre-cooled PBS and stored at −80°C.

2.8.1 Oil Red O staining

The mature adipocytes were washed three times with PBS and subjected to fixation in 4% paraformaldehyde followed by staining with 60% Oil Red O solution (Sigma).

2.9 Sequencing and data analysis

All data analysis processes were based on the self-built database MWDB (Metware Biotechnology Co., Ltd. Wuhan, China).

The intracellular lipids were extracted and analyzed by the liquid chromatography–mass spectrometry (LC–MS) system (Chen, 2009).

Consideration was given to metabolites with a variable important in projection (VIP) values of ≥1 and having fold changes ≥2 or fold changes ≤0.5. Using the R package MetaboAnalystR, VIP values were extracted from the OPLS-DA (orthogonal projections to latent structures discriminant analysis) data, which also included score plots and permutation plots.

Identified metabolites were annotated using the KEGG Compound database (http://www.kegg.jp/kegg/compound/), and annotated metabolites were then mapped to the KEGG Pathway database (http://www.kegg.jp/kegg/pathway.html). Pathways with significantly regulated metabolites were then fed into MSEA (Metabolite Set Enrichment Analysis), and their significance was determined by hypergeometric test's *p*-values.

2.10 Statistical analyses

Three independent biological experiments were conducted and the corresponding mean values were calculated. All data are reported as the mean ± SEM, and all statistical analyses were carried out using the *t*-test. Differences with a value of $p < 0.05$ were considered statistically significant.

3 Results

3.1 Cloning and validation of the ACSL1 gene

The primers were designed using the RACE primer design principle for 3′RACE PCR, and fragments with the 450 bp length were obtained using primers SP1 and SP2 (Figure 1A). The sequencing results revealed a G–A mutation site in the cloned sequence. The G–A mutation led to premature termination of the ACSL1 mRNA translation pathway. This would result in the formation of a short ACSL1 protein.

Amplification of the CDS region downstream with primers SP3 and SP4 produced distinct lengths of 3′UTR (748 bp and 1,579 bp) (Figure 1B), resulting in the CDS region having two different lengths (2,100 bp and 1,836 bp). Therefore, there are two different length CDS regions (2,100 bp and 1,836 bp) for ACSL1.

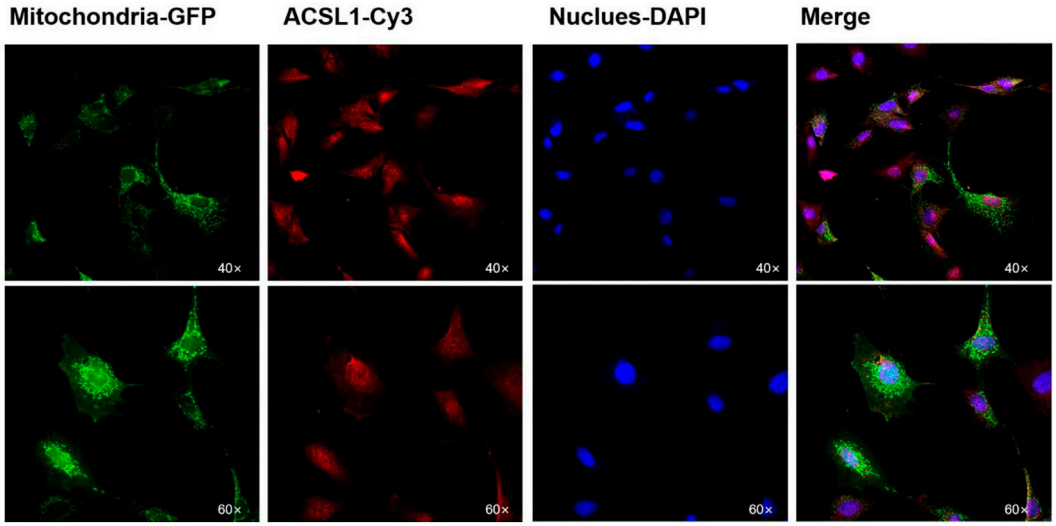


FIGURE 3
Subcellular localization of ACSL1 in sheep preadipocytes.

The results were further verified using Northern blotting which was used to detect the mRNA expression of ACSL1 gene transcripts in different tissues and stages of sheep preadipocyte differentiation (Figure 1C). The known sequence was used for the designing of primers. The 471 bp long probe was located at 320 bp in the CDS region. The results demonstrated that there are at least two different transcripts in the heart, liver, and muscle tissues. Preadipocyte differentiation in sheep involves two transcripts at different stages, and the expression of ACSL1 gene mRNA was found to be increasing during adipocyte differentiation.

The expression of two transcripts of the ACSL1 gene in sheep preadipocytes at different stages of differentiation stages was detected using Western blotting (Figure 1D). The proteins of ACSL1 have molecular weights of 78.16 and 68.06 kDa.

Northern and Western blotting results revealed the presence of at least two transcripts of the ACSL1 gene. Therefore, the transcript with a 2,100 bp long CDS region is defined as ACSL1-a, along with the other transcript ACSL1-b of 1,836 bp length. ACSL1-a has two distinct 3′UTRs. Genome sequence alignment revealed the presence of 21 exons in ACSL1-a, while ACSL1-b terminated early at exon 18 also these two transcripts share the same region before exon 18 (Figures 1E,F).

ACSL1-a protein levels gradually increased during preadipocyte differentiation, reaching their peak on the day 6. ACSL1-b protein also increased in the preadipocytes and reached the highest levels on the day 2.

3.2 Bioinformatics analysis and expression of the ACSL1 gene in sheep

Sheep ACSL1-a encodes 699 amino acids with a protein of molecular weight of 78.16 kDa; however, ACSL1-b encodes 611 amino acids with a protein having a molecular weight of 68.06 kDa. A comparison of the amino acid sequences of the transcripts revealed the absence of 88 amino acids in ACSL1-b with 86.98% similarity among amino acid sequences. ACSL1-a and ACSL1-b have protein instability coefficients of 34.25 and 34.17, respectively, and are both stable proteins. Protean was used for the prediction of secondary structures of two transcripts of the sheep ACSL1 gene. There are 32 α-helices, 47 β-corners, 35 t-corners, and 29 irregular curls in the ACSL1-b protein whereas ACSL1-b protein includes 27 α-helices, 41 β-corners, 32 t-corners, and 27 irregular curls (Figure 2A). The secondary structure folds into the tertiary structure, along with the prediction of ACSL1 protein by SWISS-MODEL is depicted in the figure (Figure 2B).

The sheep ACSL1 transcript was compared to human, mouse, pig, and cattle transcripts available on NCBI. The sheep ACSL1 gene transcript demonstrated higher homology with the cattle ACSL1 gene transcript which was consistent with the law of evolution (Figure 2C). The expression levels of different ACSL1 transcripts in 40 day-old sheep were analyzed and the results demonstrated a similarity between the expression of ACSL1-a and ACSL1-b in different tissues. The expression level of ACSL1 was the highest in adipose tissue, followed by the liver, muscle, heart, kidney, duodenum, spleen, lung, and the lowest expression level was found in the stomach (Figure 2D).

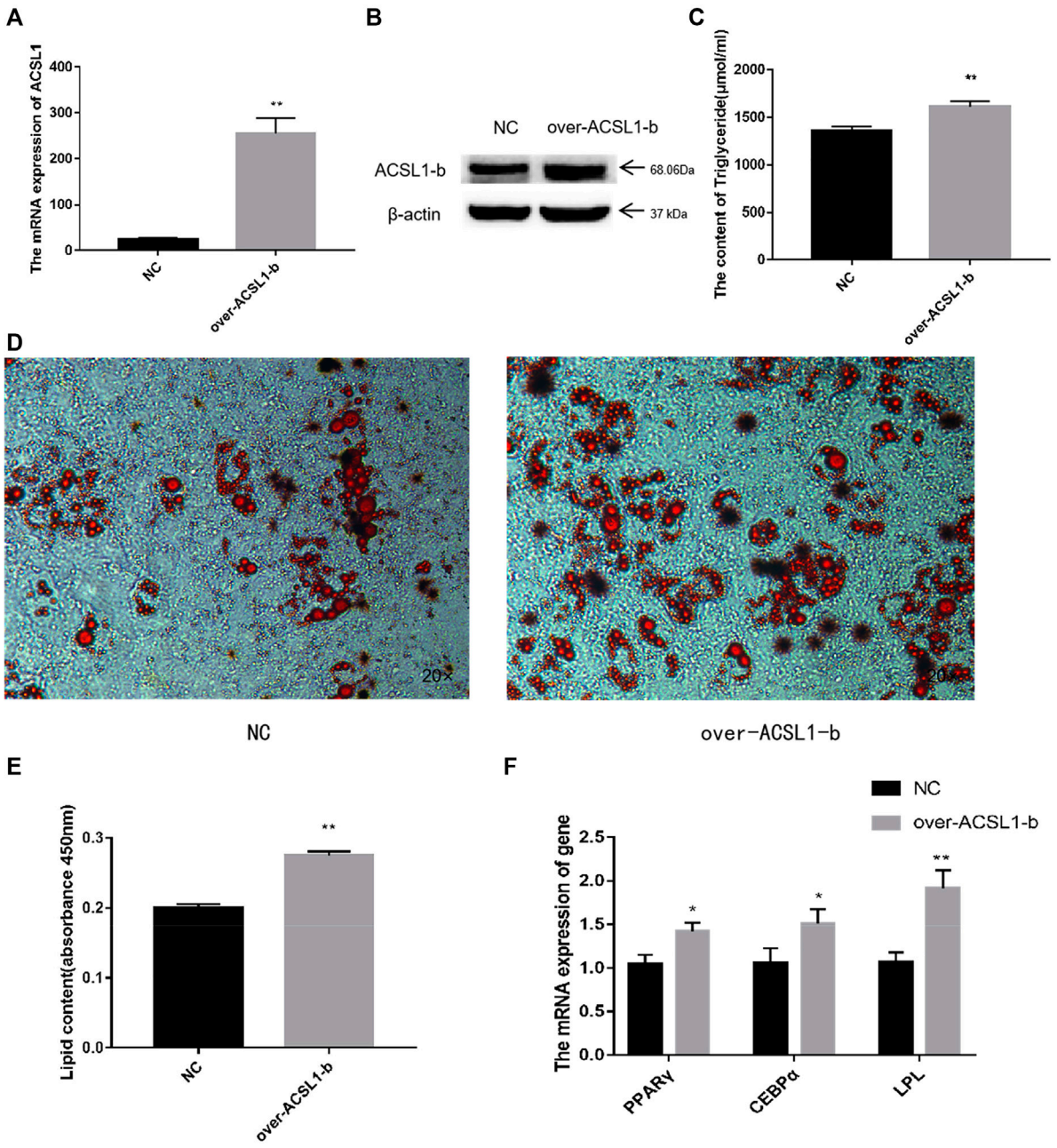


FIGURE 4
Overexpression of the ACSL1-b in sheep preadipocytes: **(A)** mRNA expression of the ACSL1-b on day 8 of differentiation (*$p < 0.05$, **$p < 0.01$). **(B)** Protein expression of the ACSL1-b on day 8 of differentiation. **(C)** Detection of intracellular triglyceride content. **(D)** Oil Red staining of adipocytes after induced differentiation on day 8 (left: control; right: overexpression ACSL1-b). **(E)** Absorbance value of Oil Red O after extraction. **(F)** mRNA expression of PPARγ, CEBPα, and LPL in sheep adipocytes. NC: adipocytes transfected with the pBI-CMV3 vector for control. Over-ACSL1: adipocytes transfected with the pBI-CMV3 vector with ACSL1-b CDS.

PSORT was used to predict the subcellular localization of the ACSL1 gene in cells, and ACSL1-a was found to be located in mitochondria, nucleus, and cytoplasm; however, ACSL1-b protein was identified in mitochondria and nucleus (Table 2). Laser confocal imaging revealed that the ACSL1 gene expression was observed in both mitochondria and nucleus, which was consistent with the online prediction (Figure 3).

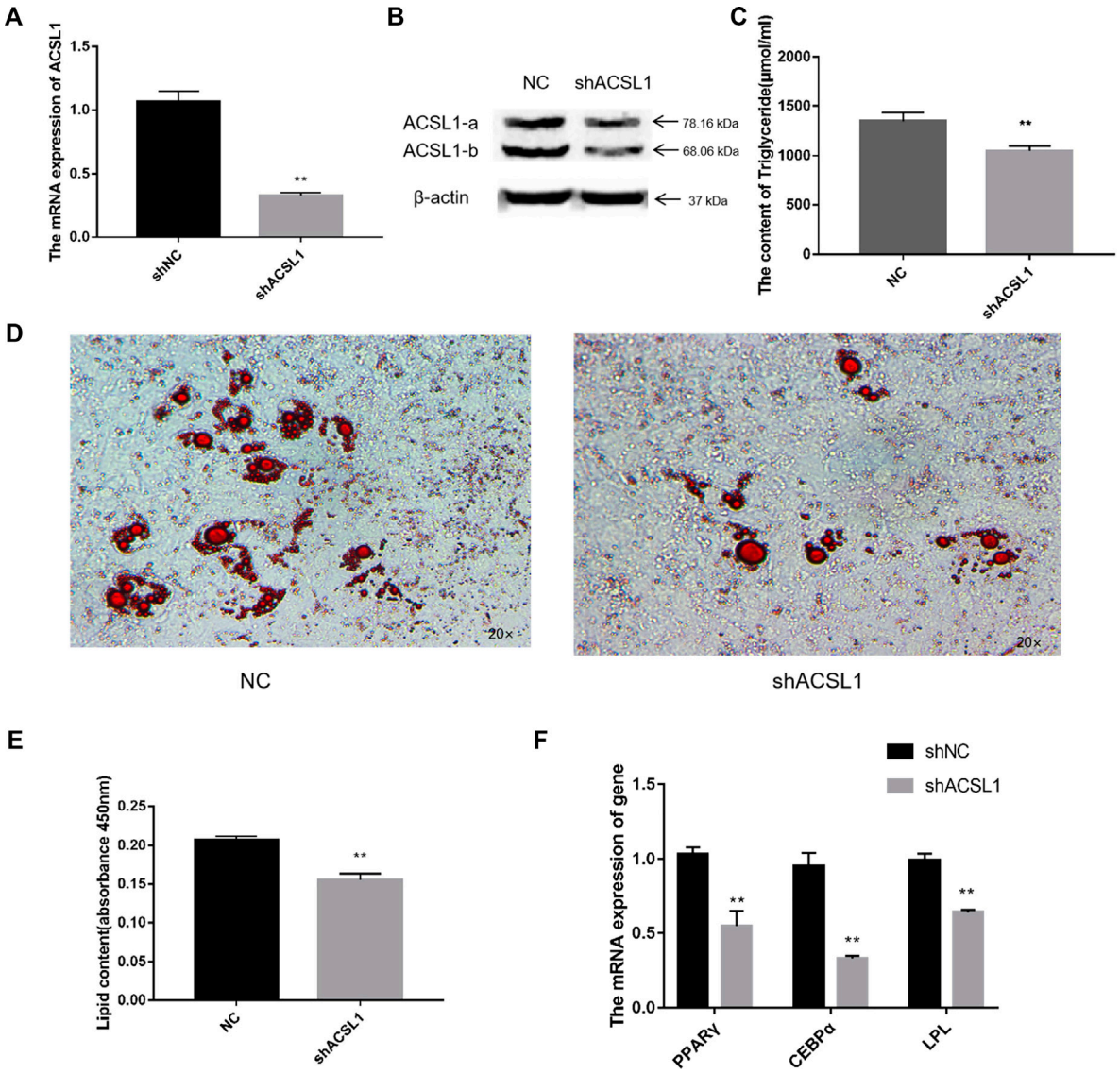


FIGURE 5
Knocked down of the ACSL1 gene in sheep preadipocytes: **(A)** mRNA expression of the ACSL1 gene on day 8 of differentiation (*$p < 0.05$, **$p < 0.01$). **(B)** Protein expression of the ACSL1 gene on day 8 of differentiation. **(C)** Detection of intracellular triglyceride content. **(D)** Oil Red staining of adipocytes after induced differentiation on day 8 (left: control; right: knocked down ACSL1). **(E)** Absorbance value of Oil Red O after extraction. **(F)** mRNA expression of PPARγ, CEBPα, and LPL in sheep adipocytes. NC: adipocytes transfected with shNC for control. ShACSL1: adipocytes transfected shRNA-297 of ACSL1 gene.

3.3 Overexpression of ACSL1-b in sheep adipocytes increased the triglyceride content

Previously, researchers from our laboratories reported that there was no substantial increase in the total triglyceride content in the cell after ACSL1-a overexpression. Lipidome sequencing, verification, and overexpression of ACSL1-a can enhance the synthesis of diglycerides as well as certain triglyceride metabolites. These results suggested that the major function of ACSL1-a might be associated with the boosting of diglycerides synthesis (Cao et al., 2020).

The CDS region of the ACSL1-b gene was cloned and ligated into the pBI-CMV3 vector, and the vectors were further transfected into sheep preadipocytes. ACSL1-b mRNA and protein expression in sheep adipocytes increased significantly on day 8 after of differentiation induction (Figures 4A,B). The transfected cells further developed into mature adipocytes.

The total TGA content analysis of adipocytes from the overexpression group was significantly higher than that of the

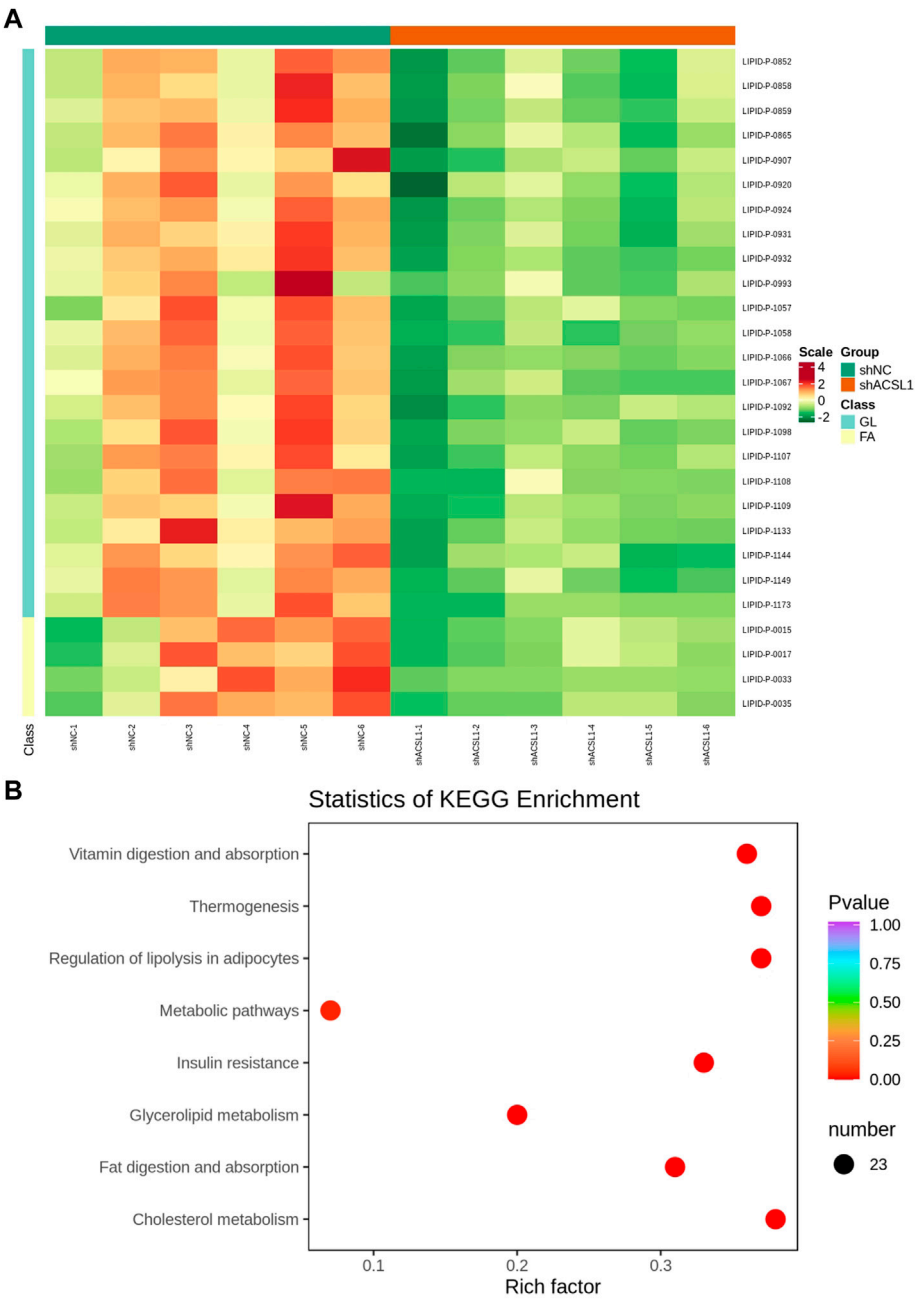


FIGURE 6
Lipid metabolome sequencing: **(A)** heat map; **(B)** KEGG analysis. shNC: adipocytes transfected with shNC as control. ShACSL1: adipocytes transfected shRNA-293 of the ACSL1 gene.

control group (Figure 4C). Furthermore, the number of lipid droplets generated in the ACSL1-b overexpression group was significantly higher than that of the control group (Figure 4D). The Oil Red O in the cells was extracted for absorbance detection. The OD value of the ACSL1-b overexpression group was significantly higher than that of the control (Figure 4E). The mRNA expression of PPARγ, CEBPα, and LPL in sheep adipocytes also increased significantly (Figure 4F).

3.4 Lipidome analysis of metabolites post ACSL1 gene knockdown

To interfere with the expression of ACSL1, shRNA targeting a common sequence using both the ACSL1 transcripts was designed. The mRNA and protein levels of both transcripts were dramatically reduced after the interference (Figures 5A,B), similarly, the triglycerides

TABLE 3 Differential metabolites in adipocytes knocking down the ACSL1 gene compared with the controls.

Compound	Class	Fold change	Type
Lauroyl-carnitine	CAR	0.381792978	Down
Myristoyl-carnitine	CAR	0.400777014	Down
Tetradecenoyl-carnitine	CAR	0.206371467	Down
Palmitoleoyl-carnitine	CAR	0.314336933	Down
TG (16:0/16:1/20:1)	TG	0.48022308	Down
TG (16:0/16:1/22:1)	TG	0.498540902	Down
TG (14:0/20:1/20:1)	TG	0.433355204	Down
TG (14:0/20:1/22:1)	TG	0.481539957	Down
TG (16:0/16:1/18:2)	TG	0.453217617	Down
TG (18:0/18:1/18:2)	TG	0.448890681	Down
TG (14:0/20:1/20:2)	TG	0.360978679	Down
TG (18:1/18:2/20:0)	TG	0.401348747	Down
TG (14:0/20:1/22:2)	TG	0.321364064	Down
TG (18:1/18:1/18:2)	TG	0.458430749	Down
TG (16:0/16:1/22:4)	TG	0.483735963	Down
TG (14:0/20:1/20:4)	TG	0.357699625	Down
TG (14:0/20:4/22:1)	TG	0.362040332	Down
TG (18:0/18:3/20:2)	TG	0.397535756	Down
TG (16:0/16:0/22:6)	TG	0.461920941	Down
TG (16:0/16:1/22:5)	TG	0.38943777	Down
TG (18:1/18:3/20:2)	TG	0.388550917	Down
TG (16:0/18:3/22:3)	TG	0.393336742	Down
TG (14:0/20:1/22:5)	TG	0.358915091	Down
TG (16:0/16:1/22:6)	TG	0.389759949	Down
TG (18:1/18:3/22:3)	TG	0.327154868	Down
TG (18:0/18:2/22:5)	TG	0.299294195	Down
TG (14:0/22:2/22:6)	TG	0.226285955	Down

and lipid droplets in the cells also reported a significant reduction (Figures 5C–E). The mRNA expression of PPARγ, CEBPα, and LPL in sheep adipocytes reduced significantly (Figure 5F). Lipidome sequencing was used in this work to detect alterations in cells.

The metabolome contained 469 metabolites, 27 of which were differential metabolites, with four acylcarnitine metabolites and 23 triglycerides metabolites shown to be downregulated (Figure 6A; Table 3). The amount of 14–22 carbon triglyceride metabolites and acylcarnitine metabolites, such as lauroyl-carnitine, myristoyl-carnitine, tetradecenoyl-carnitine, and palmitoleoyl-carnitine decreased significantly.

These metabolites are enriched in some metabolic pathways including vitamin digestion and absorption, thermogenesis, lipolysis regulation in adipocytes, metabolic pathways, insulin resistance, glycerolipid metabolism, fat digestion and adsorption, and cholesterol metabolism (Figure 6B).

4 Discussion

The transcription of DNA into mRNA is a complex process that is essential for the genetics. A gene can produce various transcripts with varying levels of expression. Various transcripts of the ACSL1 gene in humans, pigs, mice, cattle, and other organisms have been described in various investigations. To date, five transcripts of the human ACSL1 gene, eight transcripts of the pig ACSL1 gene, six transcripts of the mouse ACSL1 gene, and four transcripts of the bovine ACSL1 gene have been published in GenBank. Two transcripts of the sheep ACSL1 gene were found by RACE and verified by Northern and Western blotting. The G–A mutation site in the cloned sequence leads to the early termination of the ACSL1 mRNA translation process. After the verification of this site on the genome, it was found that there was no SNP at this site. No SNP was found at the mutation site leading to the assumption that this base substitution was caused by RNA-editing, which is epigenetic.

Mitochondria and nuclear localization of the ACSL1 gene of sheep preadipocytes were observed through confocal laser microscopy; cytoplasmic localization was predicted as well. Various reports are also available which reveal the presence of the ACSL1 gene in the endoplasmic reticulum, nucleus, plasma membrane, mitochondria, and lipid droplets in different cell lines (Gargiulo et al., 1999; Brasaemle et al., 2004). Similarly, the ACSL1 gene was found in rat adipocytes (Sleeman et al., 1998). Furthermore, the ACSL1 gene was also found in mitochondria of PtK2 epithelial cells (Milger et al., 2006), the endoplasmic reticulum, and the cytoplasm of rat liver cells (Lewin et al., 2001). Therefore, the presence of the differentially characterized ACSL1 cannot be attributed to the specific organelle.

Since its discovery, APA has been thought to be associated with to biological functions since its discovery and may be used as a molecular marker in a variety of fields in the future. The APA phenomenon of ACSL1 has been confirmed to serve a special function in the organism. Various studies have shown that the ACSL1 gene may regulate colorectal cancer through lipid metabolism. The 3′UTR diversity of the ACSL1 gene may indicate the possibility of recurrence after treatment. Patients with genotype TT are more likely to relapse than those with genotype TC (Vargas et al., 2016). Researchers have different perspectives on the ACSL1 gene studies. Various studies believe that the ACSL1 gene can promote triglyceride synthesis (Chiu et al., 2001; Pan et al., 2010). Although some studies show that the ACSL1 gene can promote the synthesis of diglycerides and glycerophospholipids, but not the synthesis of triglycerides (Mukherjee and Yun, 2012). The present study research reveals that overexpression of ACSL1-a can boost intracellular diglyceride synthesis (Cao et al., 2020) and

that the ACSL1-b gene can also promote triglyceride synthesis. The diglyceride pathway is primarily responsible for triglyceride synthesis in adipose tissue; however, there may be other interactions between the two transcripts that help to regulate triglycerides synthesis (Cao et al., 2003). In both the experiments of overexpression of ACSL1-a and knocking down of ACSL1, differential metabolites enriched in the lipolysis regulation in the adipocyte pathway were observed, which is consistent with the main function of the ACSL1 gene. Previous research has indicated that ACSL1 primarily affects the synthesis of 16–18 carbon triglycerides (Suzuki et al., 1990); however, the data show that the ACSL1 expression also influences the synthesis of 14–22 carbon triglycerides. Acylcarnitine, a byproduct of acyl-CoA oxidation, is also regulated by the ACSL1 gene. Therefore, it is believed that different transcripts of the ACSL1 gene might have different functions, but when the common region of the two transcripts knocks down the ACSL1 gene, only triglyceride synthesis is impacted, leading researchers to ignore the ACSL1 gene's role in diglyceride synthesis. The particular reasons for these phenomena require additional investigation.

5 Conclusion

Based on the length differences of the sheep ACSL1 genes, two transcripts encode separate stable proteins. ACSL1-a can promote intracellular diglyceride synthesis; moreover, ACSL1-b can promote triglyceride synthesis. Lipidome sequencing revealed that when ACSL1 transcripts were knocked down, intracellular 14–22 carbon triglycerides and acylcarnitine levels were significantly reduced. Different ACSL1 transcripts perform distinct functions in sheep lipid metabolism and the combination of ACSL1-a and ACSL1-b makes the ACSL1 gene essential for triglyceride synthesis in sheep preadipocytes.

Data availability statement

The data presented in the study are deposited in the Metabolights repository, accession number MTBLS5808.

Ethics statement

The animal study was reviewed and approved by the Animal Welfare and Ethics Committee of Jilin Academy of Agricultural Sciences. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

Data curation, YC (1st author) and YY; formal analysis, LL; funding acquisition, HJ and YL; methodology, LZ and YC (9th author); project administration, YC (1st author), YY, and YC (9th author); resources, YC (1st author); writing–original draft, YC (9th author); writing–review and editing, YY, SW, and KZ.

Funding

This research was funded by Agricultural Science and Technology Innovation Program of Jilin Province (CXGC2021ZY108), Changchun Science and Technology Development Plan (21ZY14) and China Agriculture Research System of MOF and MARA (CARS38).

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

Brasaemle, D. L., Dolios, G., Shapiro, L., and Wang, R. (2004). Proteomic analysis of proteins associated with lipid droplets of basal and lipolytically stimulated 3T3-L1 adipocytes. *J. Biol. Chem.* 279 (45), 46835–46842. doi:10.1074/jbc.M409340200

Cao, J., Lockwood, J., Burn, P., and Shi, Y. (2003). Cloning and functional characterization of a mouse intestinal acyl-CoA:monoacylglycerol acyltransferase, MGAT2 *J. Biol. Chem.* 278 (16), 13860–13866. doi:10.1074/jbc.M300139200

Cao, Y., Jin, H. G., Ma, H. H., and Zhao, Z. H. (2017). Comparative analysis on genome-wide dna methylation in longissimus dorsi muscle between small tailed han and dorper ×small tailed han crossbred sheep. *Asian-Australas. J. Anim. Sci.* 30 (11), 1529–1539. doi:10.5713/ajas.17.0154

Cao, Y., Wang, S., Liu, S., Wang, Y., Lian, Z., Ma, H., et al. (2020). Effects of long-chain fatty acyl-coa synthetase 1 on diglyceride synthesis and arachidonic acid metabolism in sheep adipocytes. *Int. J. Mol. Sci.* 21 (6), 2044. doi:10.3390/ijms21062044

Chen, Y., Zhang, R., Song, Y., He, J., Sun, J., Bai, J., et al. (2009). RRLC-MS/MS-based metabonomics combined with in-depth analysis of metabolic correlation network: Finding potential biomarkers for breast cancer *Analyst*. 134, 2003

Chikwanha, O. C., Vahmani, P., Muchenje, V., Dugan, M. E. R., and Mapiye, C. (2017). Nutritional enhancement of sheep meat fatty acid profile for human health and wellbeing. *Food Res. Int.* 104, 25–38. doi:10.1016/j.foodres.2017.05.005

Chiu, H. C., Kovacs, A., Ford, D. A., Hsu, F. F., Garcia, R., Herrero, P., et al. (2001). A novel mouse model of lipotoxic cardiomyopathy. *J. Clin. Invest.* 107 (7), 813–822. doi:10.1172/JCI10947

Crouse, J. D., Cross, H. R., and Seideman, S. C. (1984). Effects of a grass or grain diet on the quality of three beef muscles. *J. Animal Sci.* 58 (3), 619–625. doi:10.2527/jas1984.583619x

Ellis, J. M., Li, L. O., Wu, P. C., Koves, T. R., Ilkayeva, O., Stevens, R. D., et al. (2010). Adipose acyl-CoA synthetase-1 directs fatty acids toward beta-oxidation and is required for cold thermogenesis. *Cell Metab.* 12 (1), 53–64. doi:10.1016/j.cmet.2010.05.012

Gargiulo, C. E., Stuhlsatz-Krouper, S. M., and Schaffer, J. E. (1999). Localization of adipocyte long-chain fatty acyl-CoA synthetase at the plasma membrane. *J. Lipid Res.* 40 (5), 881–892. doi:10.1016/s0022-2275(20)32123-4

Gluchowski, N. L., Becuwe, M., Walther, T. C., and Farese, R. V. (2017). Lipid droplets and liver disease: From basic biology to clinical implications. *Nat. Rev. Gastroenterol. Hepatol.* 14 (6), 343–355. doi:10.1038/nrgastro.2017.32

Lewin, T. M ., Kim, J. H., Granger, D. A., Vance, J. E., and Coleman, R. A. (2001). Acyl-CoA synthetase isoforms 1, 4, and 5 are present in different subcellular membranes in rat liver and can be inhibited independently. *J. Biol. Chem.* 276 (27), 24674–24679. doi:10.1074/jbc.M102036200

Marszalek, J. R., Kitidis, C., Dararutana, A., and Lodish, H. F. (2004). Acyl-CoA synthetase 2 overexpression enhances fatty acid internalization and neurite outgrowth. *J. Biol. Chem.* 279 (23), 23882–23891. doi:10.1074/jbc.M313460200

Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138 (4), 673–684. doi:10.1016/j.cell.2009.06.016

Milger, K., Herrmann, T., Becker, C., Gotthardt, D., Zickwolf, J., Ehehalt, R., et al. (2006). Cellular uptake of fatty acids driven by the ER-localized acyl-CoA synthetase FATP4. *J. Cell Sci.* 119, 4678–4688. doi:10.1242/jcs.03280

Mukherjee, R., and Yun, J. W. (2012). Long chain acyl CoA synthetase 1 and gelsolin are oppositely regulated in adipogenesis and lipogenesis. *Biochem. Biophys. Res. Commun.* 420 (3), 588–593. doi:10.1016/j.bbrc.2012.03.038

Pan, Z., Lu, J., Lu, L., and Wang, J. (2010). Cloning of goose ACSL1 gene and its role in the formation of goose fatty liver. *J. Animal Husb. Veterinary Med.* 11, 1407–1413. doi:10.1097/MOP.0b013e3283423f35

Paul, D. S., Grevengoed, T. J., Pascual, F., Ellis, J. M., Willis, M. S., and Coleman, R. A. (2014). Deficiency of cardiac Acyl-CoA synthetase-1 induces diastolic dysfunction, but pathologic hypertrophy is reversed by rapamycin. *Biochim. Biophys. Acta* 1841 (6), 880–887. doi:10.1016/j.bbalip.2014.03.001

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320 (5883), 1643–1647. doi:10.1126/science.1155390

Sleeman, M. W ., Donegan, N. P., HelleR-HaRRison, R., Lane, W. S., and Czech, M. P. (1998). Association of acyl-CoA synthetase-1 with GLUT4-containing vesicles. *J. Biol. Chem.* 273 (6), 3132–3135. doi:10.1074/jbc.273.6.3132

Souza, S. C., Muliro, K. V., Liscum, L., Lien, P., Yamamoto, M. T., Schaffer, J. E., et al. (2002). Modulation of hormone-sensitive lipase and protein kinase A-mediated lipolysis by perilipin A in an adenoviral reconstituted system. *J. Biol. Chem.* 277 (10), 8267–8272. doi:10.1074/jbc.M108329200

Sun, Y., Fu, Y., Li, Y., and Xu, A. (2012). Genome-wide alternative polyadenylation in animals: Insights from high-throughput technologies. *J. Mol. Cell Biol.* 4 (6), 352–361. doi:10.1093/jmcb/mjs041

Suzuki, H., Kawarabayasi, Y., Kondo, J., Abe, T., NishiKawa, K., Kimura, S., et al. (1990). Structure and regulation of rat long-chain acyl-CoA synthetase. *J. Biol. Chem.* 265 (15), 8681–8685. doi:10.1016/s0021-9258(19)38942-2

Suzuki, H., Watanabe, M., Fujino, T., and YamamoTo, T. (1995). Multiple promoters in rat acyl-CoA synthetase gene mediate differential expression of multiple transcripts with 5'-end heterogeneity. *J. Biol. Chem.* 270 (16), 9676–9682. doi:10.1074/jbc.270.16.9676

Tian, B., Hu, J., Zhang, H., and Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* 33 (1), 201–212. doi:10.1093/nar/gki158

Vargas, T., Moreno-Rubio, J., Herranz, J., Cejas, P., Molina, S., Mendiola, M., et al. (2016). 3'UTR polymorphism in ACSL1 gene correlates with expression levels and poor clinical outcome in colon cancer patients. *PLoS One* 11 (12), e0168423. doi:10.1371/journal.pone.0168423

Yue, B. (2013). *Correlation analysis between ACSL1 gene polymorphism and liver performance in Landes goose*. Shihezi: Shihezi University.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1071873

OPEN ACCESS

EDITED BY
Zebin Zhang,
Uppsala University, Sweden

REVIEWED BY
Lichun He,
Nantong University, China
Bo Wang,
China Agricultural University, China

*CORRESPONDENCE
Yongqing Jiang,
jyq61@sohu.com
Xin Huang,
hxin500@163.com
Junfang Jiang,
jiangjunfang1031@sina.com

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

RECEIVED 17 October 2022
ACCEPTED 14 December 2022
PUBLISHED 04 January 2023

CITATION
Zheng K, Wu J, Ullah S, Cao Y, Jiang Y, Huang X and Jiang J (2023), Proteome changes of dairy calves rumen epithelium from birth to postweaning. *Front. Genet.* 13:1071873. doi: 10.3389/fgene.2022.1071873

Proteome changes of dairy calves rumen epithelium from birth to postweaning

Kaizhi Zheng[1], Jianliang Wu[1], Saif Ullah[2], Yang Cao[1], Yongqing Jiang[1]*, Xin Huang[1]* and Junfang Jiang[1]*

[1]Institute of Animal Husbandry and Veterinary, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]Faculty of Veterinary and Animal Sciences, Lasbela University of Agriculture Water and Marine Sciences, Lasbela, Pakistan

Background: Rumen epithelium plays a central role in absorbing, transporting, and metabolizing of short-chain fatty acids. For dairy calves, the growth of rumen papillae greatly enhances the rumen surface area to absorb nutrients. However, the molecular mechanism underlying dairy calves rumen postnatal development remains rarely understood.

Results: Here, we firstly describe the histological change of rumen epithelium from birth to day 90 of age. Then, a shotgun approach and bioinformatics analyses were used to investigate and compare proteomic profiles of Holstein calve rumen epithelium on day 0, 30, 60 and 90 of age. A total of 4372 proteins were identified, in which we found 852, 342, 164 and 95 differentially expressed proteins between D0 and D30, between D30 and D60, between D60 and D90, respectively. Finally, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed to provide a comprehensive proteomic landscape of dairy calves rumen development at tissue level.

Conclusion: To conclude, our data indicated that keratinocyte differentiation, mitochondrion formation, the establishment of urea transport and innate immune system play central roles during rumen epithelium development. Tetrahydrobiopterin (BH4) presents an important role in rumen epithelial keratinization. The biological processes of BH4 biosynthesis and molecular function of nicotinamide adenine dinucleotide phosphate binding participate in mitochondrial cristae formation. The proposed datasets provide a useful basis for future studies to better comprehend dairy calves rumen epithelial development.

KEYWORDS

proteome, dairy calves, rumen, epithelium, postnatal development

1 Introduction

Ruminants possess remarkable and distinct multi-chambered stomachs that allows them to utilize plant fiber as energy source. The consumed plant fiber is degraded by diverse microbiota in the rumen into precursors of essential metabolites and metabolic cofactors (Xu et al., 2021). For dairy calves, replacement heifers are the second largest annual operating expense on the farm, which is critical for the future of dairy farms in improving genetic merit and maintain herd size (Tozer and Heinrichs, 2001). During replacement heifer stage, the growth of rumen papillae greatly enhances the rumen surface area to absorb nutrients, while the development of the rumen epithelium is a critical physiological challenge to transit from solid diet to milk (Baldwin and Connor, 2017; Diao et al., 2019). Rumen epithelium plays a central role in absorbing, transporting, and metabolizing of short-chain fatty acids (SCFAs) after the rumen microbes digest fiber (Ji et al., 2021). SCFAs can provide up to 70% of the energy requirements of ruminants (Bergman, 1990). Thus, a well-developed rumen is essential for the raise of replacement heifer, which will influence the performance of dairy cow in future.

Unlike the monogastric stomachs, the rumen tissue owns a complex structure consisting of stratified squamous epithelium, lamina propria, mucosa, tunica muscularis and plasma membrane (Baldwin, 1998). Thus, the postnatal development of rumen epithelium differs from that of monogastric animals. After birth, dramatic structural and functional changes occur during postnatal rumen development, in which rumen papillae grow and the SCFAs metabolism machinery gradually matures (Xu et al., 2022). However, current understanding to the biological foundation of rumen epithelium development remains limited. Hence, an overview profile of protein expression during rumen epithelium development is essential to understand its biological foundation (Finnegan et al., 2019). Nowadays, proteomics, as a powerful tool, is widely applied to identify and quantify overall proteins and present content of a cell, tissue or an organism (Aslam et al., 2017). Given the central role of rumen epithelium in dairy calves production, we were promoted to use proteomic as a powerful tool to illuminate the underlying molecular mechanisms that involved in rumen epithelium postnatal development.

2 Material and methods

2.1 Animals

The male Holstein calves at the age of day 0, day 30, day 60 and day 90 were randomly selected with 5 calves in each age. Rumen epithelia were isolated and frozen in liquid nitrogen. Experimental protocols for animal research were approved by the Institutional Animal Care and Use Committees at the Zhejiang Academy of Agricultural Sciences.

2.2 Histology analysis

The histological analysis was performed essentially as the procedure described previously (Zheng et al., 2021), the rumen tissues kept in PFA were dehydrated with alcohol and embedded in paraffin, sectioned at 5 μm, and stained with hematoxylin and eosin (H&E), and the sections were observed under a microscope (Nikon, NY, United States).

2.3 Sample preparation and trypsin digestion for label-free proteome

The sample preparation of protein was performed essentially as the procedure described previously (Zheng et al., 2022). Rumen epithelium samples were ground individually in liquid nitrogen and lysed with PASP lysis buffer (100 mM NH_4HCO_3, 8 M Urea, pH 8.0), followed by 5 min of ultrasonication on ice. The lysate was centrifuged at 12000 g for 15 min at 4°C. The supernatants were collected and quantified by BCA Protein Assay Kit (Beyotime Institute of Biotechnology, Shanghai, China). 20 μg of the protein sample was loaded to 12% SDS-PAGE gel electrophoresis for quality control and the supernatant was reduced with 10 mM DTT for 1 h at 56°C, and subsequently alkylated with sufficient iodoacetamide for 1 h at room temperature in the dark. Then samples were completely mixed with 4 times volume of precooled acetone by vortexing and incubated at −20°C for at least 2 h. Samples were then centrifuged at 12000 g for 15 min at 4°C and the precipitation was collected.

After washing with 1 ml cold acetone, the pellet was dissolved by dissolution buffer (8 M Urea, 100 mM TEAB, pH 8.5). Trypsin and 100 mM TEAB buffer were added, sample was mixed and digested at 37°C for 4 h. Then trypsin and $CaCl_2$ were added digested overnight. Formic acid was mixed with digested sample, adjusted pH under 3, and centrifuged at 12000 g for 5 min at room temperature. The supernatant was slowly loaded to the C18 desalting column, washed with washing buffer (0.1% formic acid, 3% acetonitrile) 3 times, then added elution buffer (0.1% formic acid, 70% acetonitrile). The eluents of each sample were collected and lyophilized.

2.4 LC-MS/MS analysis

LC-MS/MS analysis was then performed according to the procedure described previously (Zhou C. Y. et al., 2019). Mobile phase A (0.1% formic acid in H_2O) and B solution (0.1% formic acid in 80% acetonitrile) were prepared. The lyophilized powder was dissolved in 10 μL of solution A, centrifuged at 14000 g for

20 min at 4°C, and 1 μg of the supernatant was injected into a home-made C18 Nano-Trap column (4.5 cm × 75 μm, 3 μm). Peptides were separated in a home-made analytical column (15 cm × 150 μm, 1.9 μm), using a linear gradient elution.

LC-MS/MS analyses were performed by using Q ExactiveTM series mass spectrometer (Thermo Fisher, Germany) at Novogene Co., Ltd. (Beijing, China), with ion source of Nanospray Flex™ (ESI), spray voltage of 2.1 kV and ion transport capillary temperature of 320°C. Full scan range from m/z 350 to 1500 with resolution of 60000 (at m/z 200), an automatic gain control target value was 3×10^6 and a maximum ion injection time was 20 ms. The top 40 precursors of the highest abundant in the full scan were selected and fragmented by higher energy collisional dissociation and analyzed in MS/MS, where resolution was 15000 (at m/z 200), the automatic gain control target value was 1×10^5, the maximum ion injection time was 45 ms, a normalized collision energy was set as 27%, an intensity threshold was 2.2×10^4, and the dynamic exclusion parameter was 20 s. The raw data of MS detection was named as ".raw".

2.5 Protein identification and quantitation

The all resulting spectra were searched by Proteome Discoverer 2.2 (PD 2.2, Thermo). The corresponding proteins were matched with *bovine* database. The search parameters are set as follows: mass tolerance for precursor ion was 10 ppm and mass tolerance for product ion was 0.02 Da. Carbamidomethyl was specified as fixed modifications, Oxidation of methionine was specified as dynamic modification, and acetylation was specified as N-Terminal modification. A maximum of two missed cleavage sites were allowed. To improve the quality of analysis results, the software PD 2.2 further filtered the retrieval results: Peptide Spectrum Matches (PSMs) with a credibility of more than 99% was identified PSMs. The identified protein contains at least 1 unique peptide. The identified PSMs and protein were retained and performed with FDR no more than 1.0%. Principal component analysis (PCA) and volcano plot, which combined fold-change analysis and t-tests, were performed. Proteins with a minimum fold change of 2 (ratio >2 or <0.5, $p < 0.05$) were considered to be regulated differently between comparisons.

2.6 Bioinformatics analysis

GO functional analysis was conducted using the interproscan program against the non-redundant protein database (including Pfam, PRINTS, ProDom, SMART, ProSite, PANTHER), and the databases of COG (Clusters of Orthologous Groups) and KEGG (http://www.genome.jp/kegg/) were used to analyze the protein family and pathway. DPEs were used for Volcanic map analysis, cluster heat map analysis and enrichment analysis of GO, subcellular localization and KEGG. The probable protein-protein interactions were predicted using the STRING-db server (http://string.embl.de/).

3 Results

3.1 The histological changes of ruminal papillae during postnatal stage

To have an overview of rumen epithelium postnatal development, we firstly evaluated the histological changes of rumen papillae from birth to 90 days old. The rumen epithelium experienced dramatic transformation to tongue shape papillae on day 30 of age. Thereafter, the papilla underwent a period of growth until 90 days of age.

On the first day of Holstein diary calve postnatal life, the papillae were formed by a single layer of stratum basale cells that covered papillary core. The papillary core was covered by a few layers of stratified squamous cells, which were covered by serosa. The rumen epithelia then formed tongue-shaped papillae on day 30 of age, which were constituted by stratum basale, stratum spinosum and stratum granulosum. On 60 days of age, the space between papillae was larger than that of day 30, and stratum corneum could already be clearly observed. We could also found branch like morphology on tongue-shaped papillae since this stage. On 90 days of age, the stratum corneum covered most of the papillae except the top of papillae. In spite of more layers of stratified squamous cells, we didn't found significant histological changes of ruminal papillae on 90 days of age, when comparing with day 60 (Figure 1).

3.2 Protein identification and comparison analysis

To explore the mechanism driving Holstein calves rumen papillae development, label-free proteomic strategy was used to identify different abundant proteins of day 0, day 30, day 60 and day 90. A total of 4372 proteins were identified, including 3951 shared proteins, 31 proteins uniquely expressed in day 0, 17 proteins uniquely expressed in day 30, 5 proteins uniquely expressed in day 60 and day 90 (Figure 2A). PCA analysis was performed to visually differentiate the sample clusters among the observations. We found that 68.30% of the variability was explained by the first two principal components, which accounted for 54.84%, and 13.46% of the total variance. The rumen epithelium of day 0, day 30 and day 90 could be separated completely by identified proteins and distributed in different locations. However, rumen epithelium of day 60 could not be completely separated from that of day 30 and day 90 (Figure 2B).

The differentially abundant proteins were highlighted by simultaneously considering fold change >2 or <0.5 and $p <$

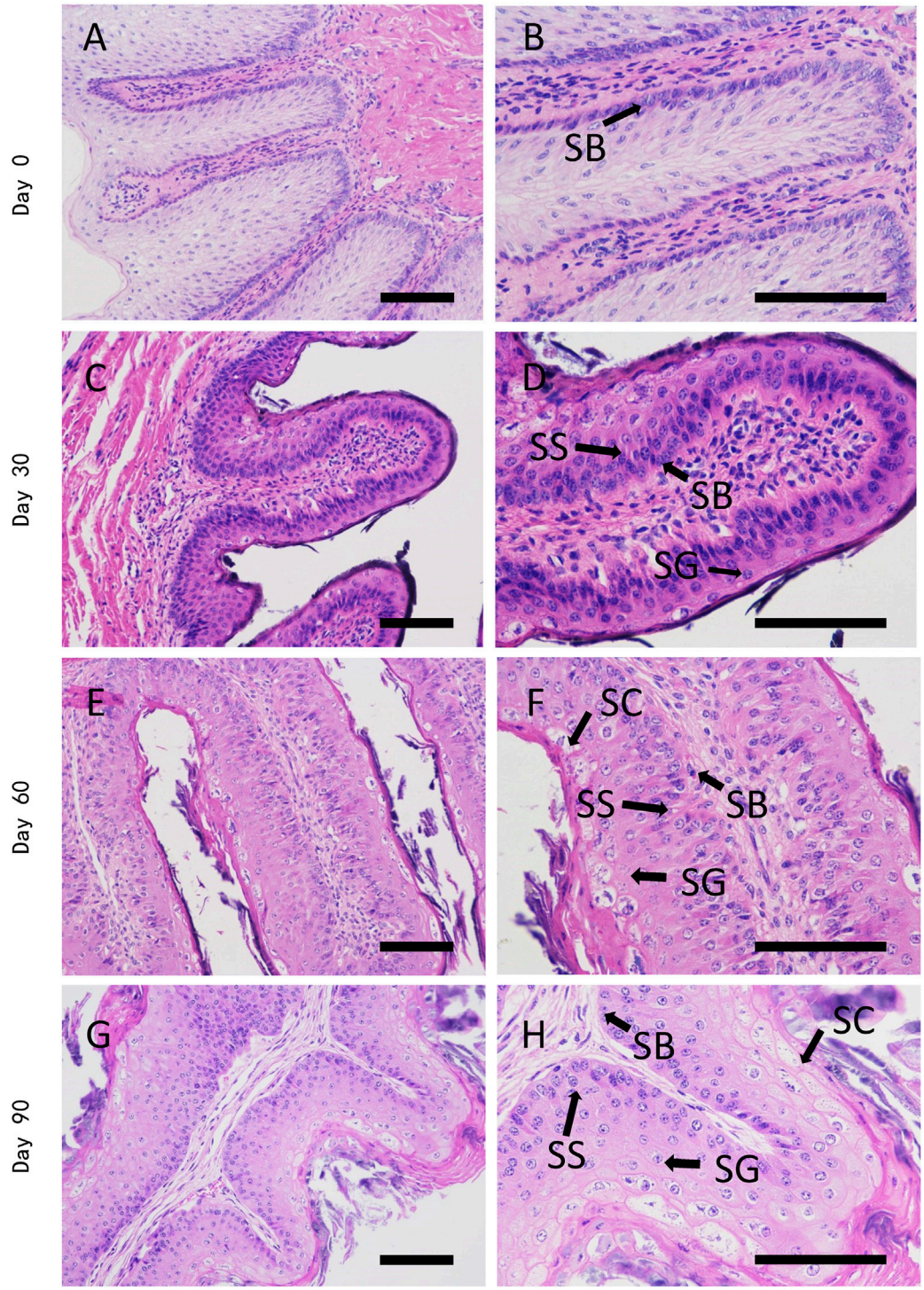


FIGURE 1
H&E staining of dairy cattle rumen epithelium on the age of day 0 **(A,B)**, day 30 **(C,D)**, day 60 **(E,F)**, and day 90 **(G,H)**. SB is stratum basale, SS is stratum spinosum, SG is stratum granulosum, SC is stratum corneum. Bar is 50 μm.

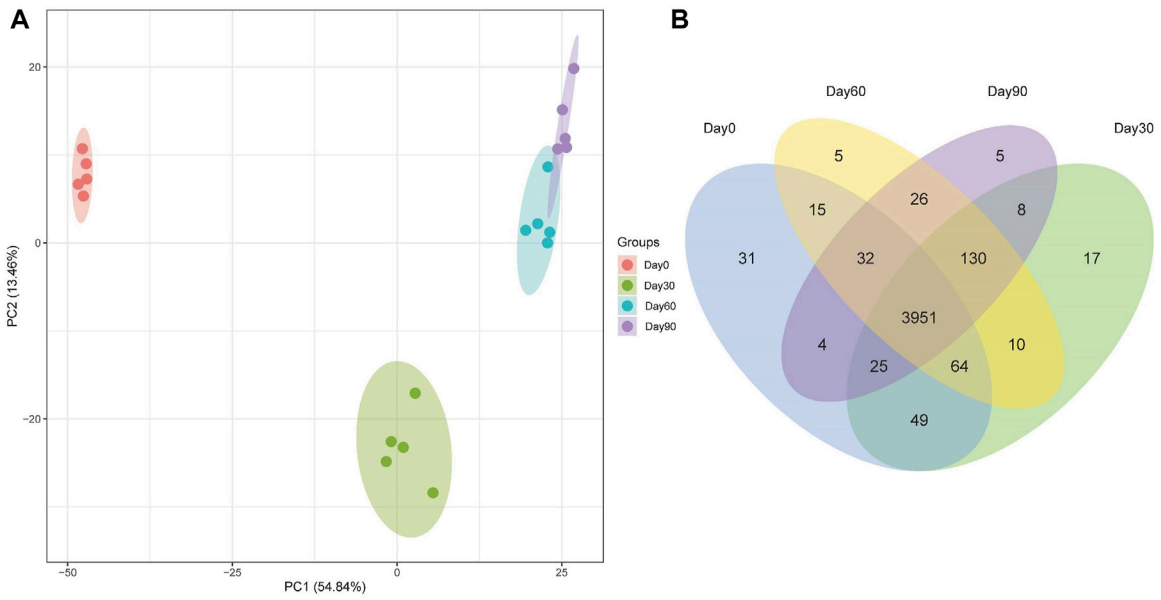


FIGURE 2
PCA scores plot **(A)** and Venn Diagram **(B)** of proteins identified in rumen epithelium at the age of day 0, day 30, day 60, and day 90.

0.05. A total of 452 proteins were significantly different in comparison of day 0 and day 30 rumen epithelium (Supplementary Table S1). While the number of DEPs was 152 between rumen epithelium of day 30 and day 60 (Supplementary Table S2). And finally, this number became 34 between day 60 and day 90 (Supplementary Table S3). Unsupervised hierarchical clustering of different biological data sets was then conducted as a more rigorous test for the screened proteins to evaluate the rationality, accuracy and dynamic changes of these DEPs (Figure 3).

3.3 Bioinformatics analysis of differentially abundant proteins

In order to assess the major biological processes that participated in the postnatal development of calves rumen epithelia, a GO annotation was performed to analyze the functional characteristics of proteins identified in calves rumen epithelia of day 0, day 30, day 60 and day 90 (Figure 4). Bioinformatics analyses were then performed to construct a specific molecular network to explore the biological functions and pathways related to the DEPs from birth to day 90 of age. GO enrichment analysis showed that DEPs in rumen epithelia were significantly enriched in 41 GO terms between day 0 and day 30 ($p < 0.05$). In the biological processes analysis, the top five biological progresses were tetrahydrobiopterin biosynthetic process (GO:0006729), cristae formation (GO:0042407), pigment biosynthetic process (GO:0046148), peptide cross-linking (GO:0018149) and immune response (GO:0006955). In molecular functions, the top five significant GO terms were NADP binding (GO:0050661), structural molecule activity (GO:0005198), N,N-dimethylaniline monooxygenase activity (GO:0004499), peptidase inhibitor activity (GO:0030414) and hydro-lyase activity (GO:0016836). As for cell components, extracellular region (GO:0005576), intermediate filament (GO:0005882), cytoskeletal part (GO:0044430), extracellular region part (GO:0044421) and cytoskeleton (GO:0005856) were the top five GO terms that enriched significantly ($p < 0.05$) (Figure 5A, Supplementary Table S4).

While the number of GO enriched terms decreased to 23 between day 30 and day 60 of age. The top five molecular function GO terms were nucleosomal DNA binding (GO:0031492), extracellular matrix structural constituent (GO:0005201), calcium ion binding (GO:0005509), regulatory region DNA binding (GO:0000975) and serine-type endopeptidase activity (GO:0004252). As for biology processes, superoxide metabolic process (GO:0006801), regulation of apoptotic process (GO:0042981), peptide cross-linking (GO:0018149), aromatic amino acid family metabolic process (GO:0009072) and glutamine biosynthetic process (GO:0006542) were the top five significantly enriched GO terms. And for cell components, the significantly enriched GO term was chromatin (GO:0000785) ($p < 0.05$) (Figure 5B, Supplementary Table S5).

Between rumen epithelium tissues of day 60 and day 90, the significantly enriched GO terms of biology progresses were

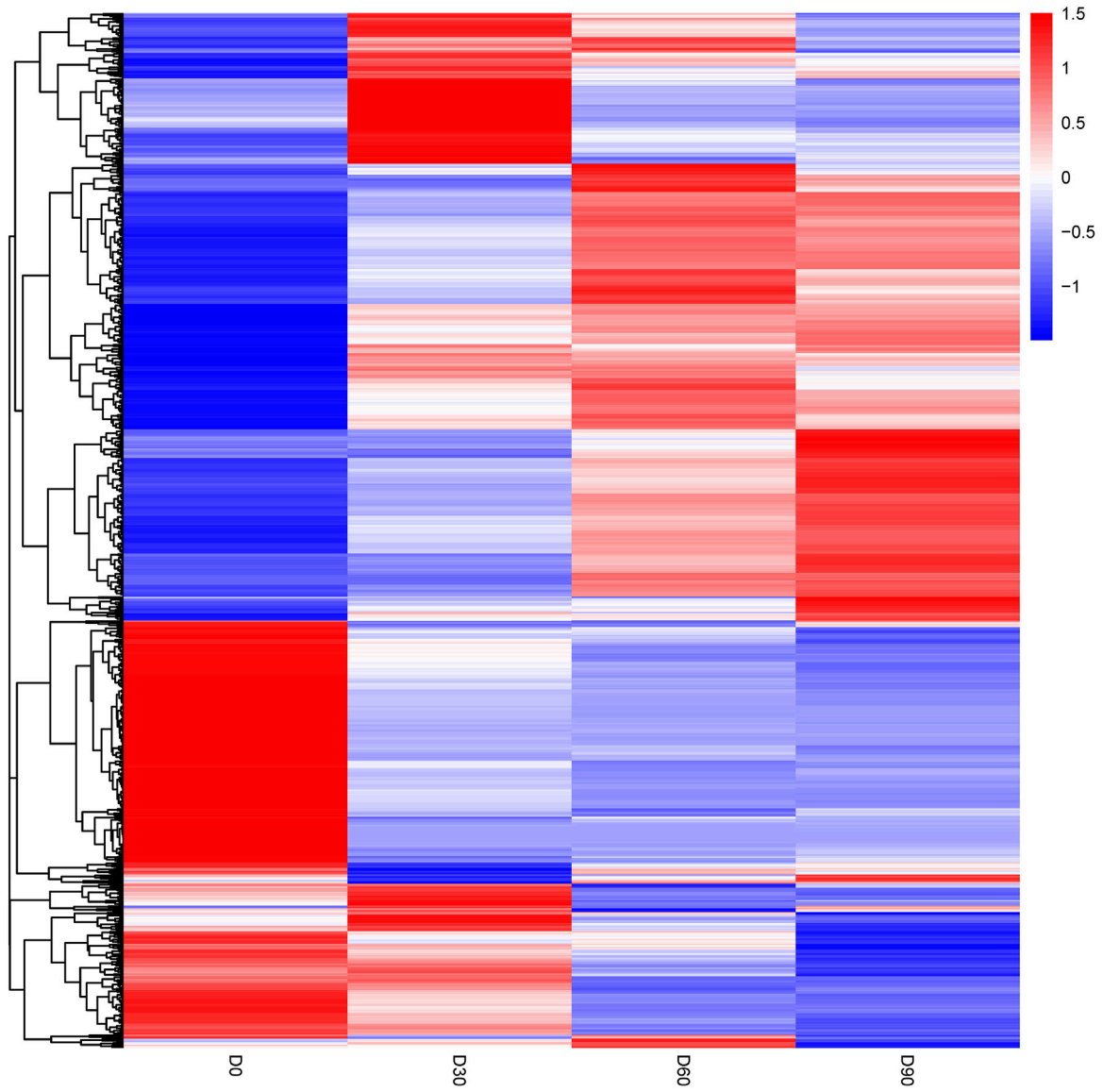


FIGURE 3
Heat map of differentially abundant proteins in rumen epithelium at the age of day 0, day 30, day 60 and day 90.

antigen processing and presentation (GO:0019882), immune response (GO:0006955), protein citrullination (GO:0018101), transcription initiation from RNA polymerase II promoter (GO:0006367), cellular protein modification process (GO:0006464), macromolecule metabolic process (GO:0043170). As for molecular functions, the GO terms included protein-arginine deiminase activity (GO:0004668), 3-hydroxyacyl-CoA dehydrogenase activity (GO:0003857), lipid binding (GO:0008289) and protein-glutamine gamma-glutamyltransferase activity (GO:0003810). In cell components, DEPs were significantly enriched in MHC class II protein complex (GO:0042613), transcription factor TFIIF complex (GO:0005674), nuclear part (GO:0044428) spliceosomal complex (GO:0005681) and protein complex (GO:0043234) ($p < 0.05$) (Figure 5C, Supplementary Table S6).

3.4 KEGG pathway enrichment analysis

Then, we performed KEGG pathway enrichment analysis to extract the biological pathways related to the DEPs in rumen epithelia from birth to day 90 of age. Between day 0 and day 30, the DEPs were significantly enriched in 24 pathways. The top five significantly enriched KEGG terms were Arachidonic acid

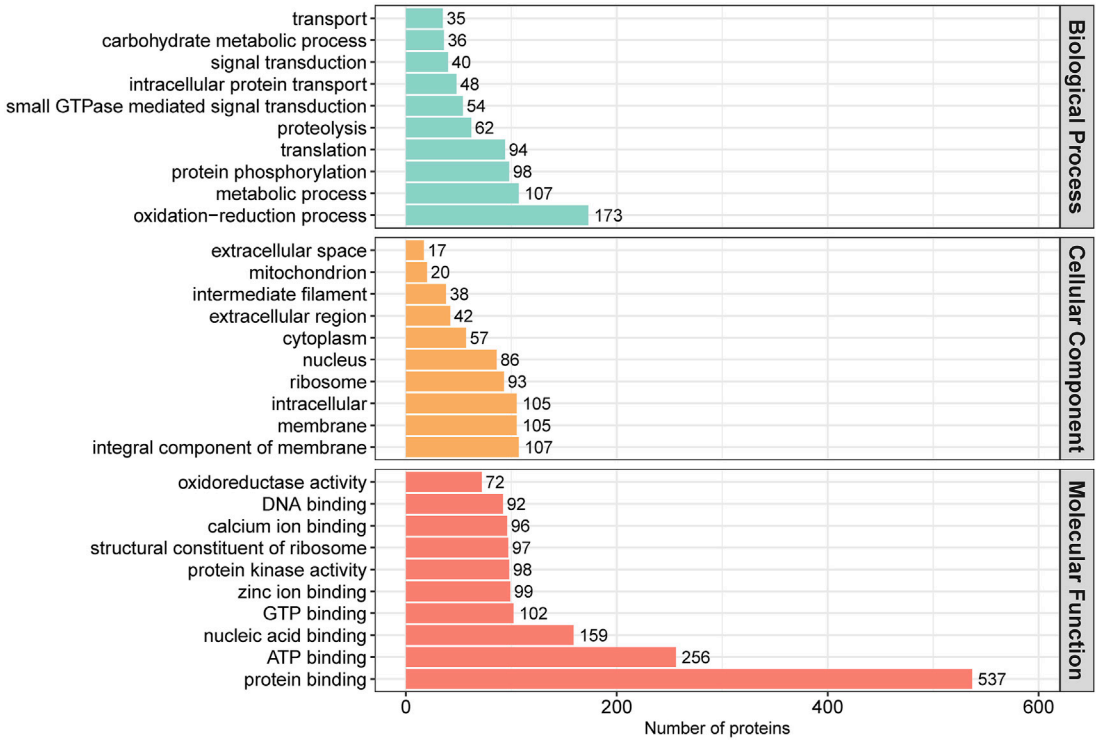


FIGURE 4
GO annotation classification of differentially abundant proteins in rumen epithelium at the age of day 0, day 30, day 60 and day 90.

metabolism (map00590), Folate biosynthesis (map00790), Steroid hormone biosynthesis (map00140), Vitamin digestion and absorption (map04977), Drug metabolism - cytochrome P450 (map00982) ($p < 0.05$) (Figure 6A, Supplementary Table S7). From day 30 to day 60 of age, the top five significantly enriched KEGG terms were NOD-like receptor signaling pathway (map04621), Salivary secretion (map04970), Nitrogen metabolism (map00910), Cytosolic DNA-sensing pathway (map04623) and Relaxin signaling pathway (map04926) ($p < 0.05$) (Figure 6B, Supplementary Table S8). From day 60 to day 90 of age, the top five significantly enriched KEGG terms were Graft-versus-host disease (map05332), Type I diabetes mellitus (map04940), Autoimmune thyroid disease (map05320), Allograft rejection (map05330), HTLV-I infection (map05166) ($p < 0.05$) (Figure 6C, Supplementary Table S9).

4 Discussion

To the best of our knowledge, the present study is the first to systemically characterize proteome changes of diary calve rumen epithelium from birth to day 90 of age. Our research also provides extensive and functional analyses of diary calve rumen epithelium, which provides a useful basis to better comprehend cow rumen epithelial development. Rumen is an essential organ in ruminants, which plays principal roles in nutrient metabolism and transportation. The development of rumen epithelium to tongue-shaped papillae increases the surface area of rumen wall, which contributes to the absorption of SCFAs (Nishihara et al., 2019). During dairy cow postpartum period, a well-developed papilla can promote its absorption rate, which increases the dry matter intake and milk production (Miller et al., 2021). Hence, the development of rumen epithelium at early stage determines the future production of milk. From monogastric animal to a ruminant, the dairy calves undergoes a series of dramatic gastrointestinal transformations after birth, which changes from a milk-based to a solid-based diet (Steele et al., 2016; Meale et al., 2017). Thus, understanding the mechanism that drives the transformation of cow rumen epithelium is important for devising strategies to improve ruminant productivity (Roh et al., 2016). However, the molecular mechanism regulating the growth of rumen epithelium remains unclear. In this study, the histological and morphological changes of calves rumen epithelium were described. We also analyzed the proteomic changes of 20 rumen tissue samples during four stages of dairy calves, which provided a comprehensive landscape dairy calves rumen epithelium from birth to day 90 of age.

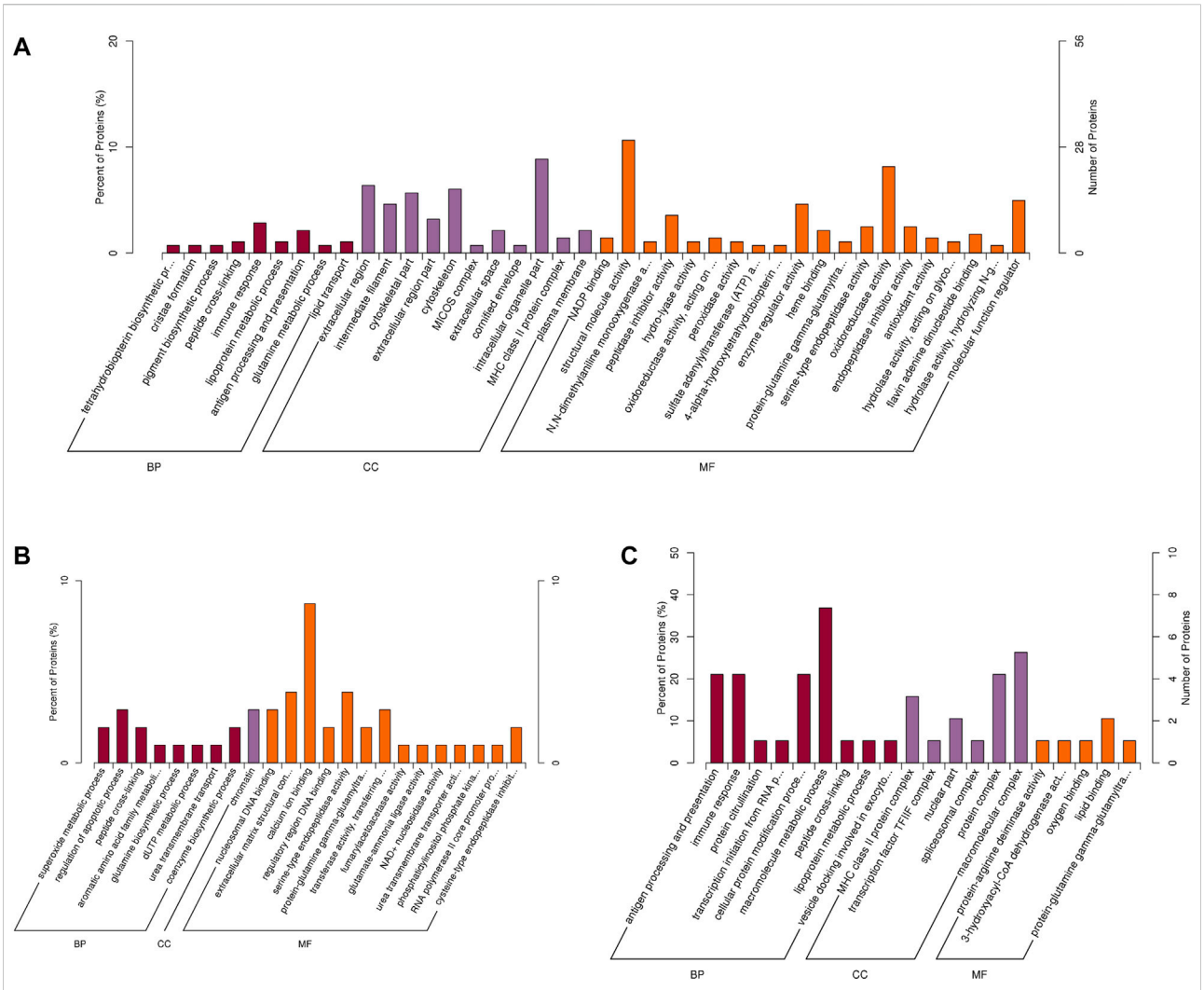


FIGURE 5
GO enrichment analysis of differentially abundant proteins in rumen epithelium between **(A)** day 0 and day 30, **(B)** day 30 and day 60, and **(C)** day 60 and day 90 of age. All pathways are listed according to *p*-value.

The high consistence of PCA and heat map analysis results showed that the DEPs upon different ages were distinguishable among rumen epithelium samples of day 0, day 30 and day 60. While the results of day 60 could not be distinguished from that of day 90. These data indicate that significant histological changes occur from day 0 to day 30 and from day 30 to day 90.

The rumen epithelium experienced a dramatic transformation from a single layer of stratum basale cells covered by a few layers of stratified squamous cells to tongue-shaped papilla constituted by stratum basale, stratum spinosum and stratum granulosum. Stratum corneum is the outermost barrier of rumen epithelium that prevent the infection of rumen microorganisms. The keratinization of stratified squamous epithelium of the rumen is similar to that of skin epidermis (Graham and Simmons, 2005). Located at the rumen epithelial surface, keratinocytes are constantly exposed to external stimuli and are the first barrier to invading pathogens (Coulombe, 2017). In keratinocytes, keratins are the major structural intermediate filament proteins, which are expressed in a highly specific pattern at different differentiation stages of keratinocytes. Thus, it is widely used as biomarkers for identifying the stage of epithelial cell proliferation or differentiation (Zhang et al., 2019). Previous studies have suggested that the KRT4 gene and keratinocyte differentiation pathway play key roles in stratum corneum formation by transcriptomic analyses (Bragulla and Homberger, 2009). A recent study has also identified KRT17 and KRT15 as main keratin-encoding gene in two different stages of rumen epithelium keratinization in sheep (Yuan et al., 2022). Consistent with previous studies, we found the protein expression of keratin 17 down-regulated on

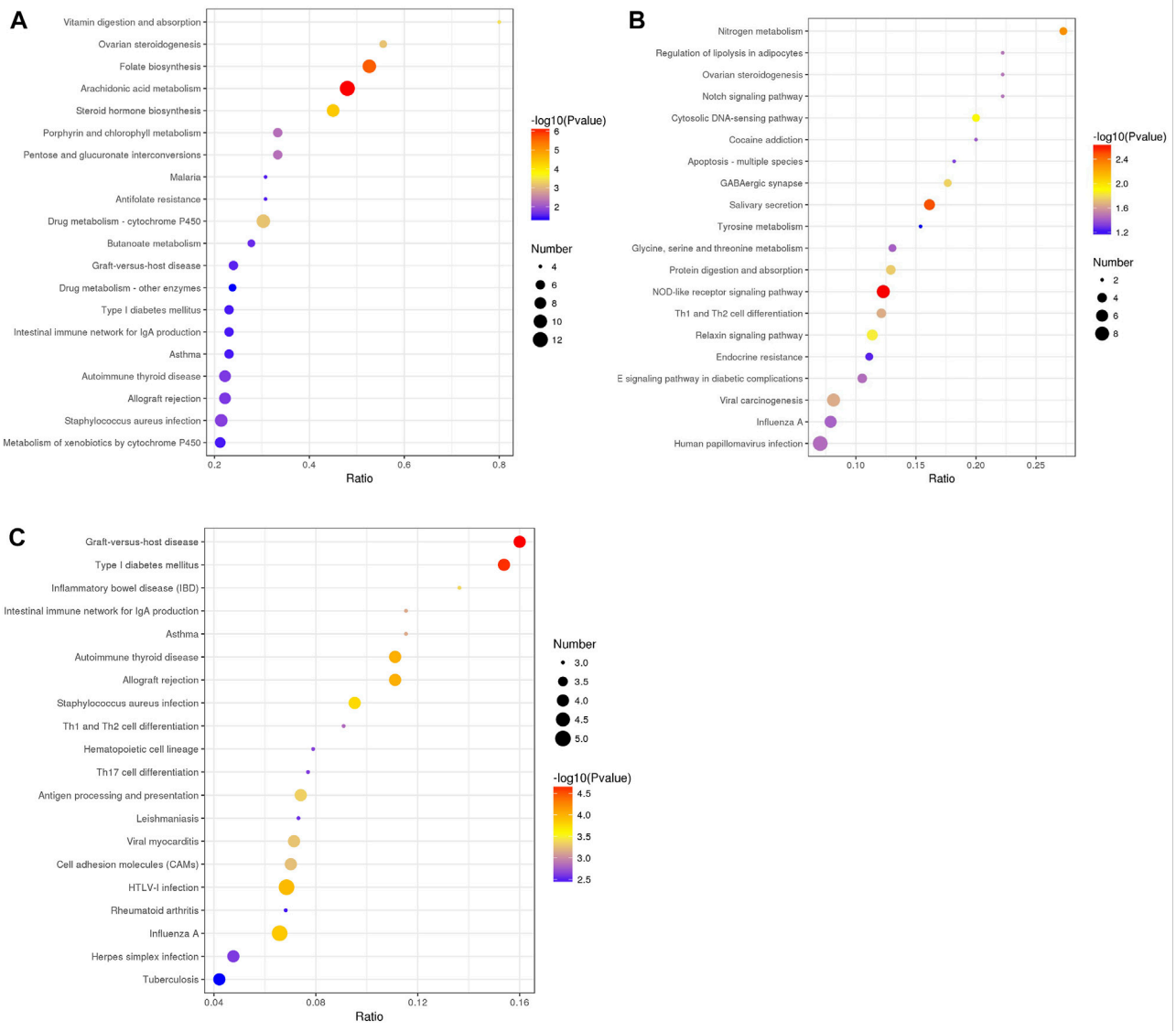


FIGURE 6
KEGG pathway enrichment analysis of differentially abundant proteins in rumen epithelium between **(A)** day 0 and day 30, **(B)** day 30 and day 60, and **(C)** day 60 and day 90 of age. All pathways are listed according to *p*-value.

day 0 of age, while the expression of keratin 15 and keratin 4 gradually decreased from day 0 to day 60 of age. These results indicated that keratinocyte differentiation gradually decreased from day 0 to day 60 of age. We also found keratin 6 up-regulated and keratin 7, keratin 80, keratin 85, keratin 75 and keratin 72 down-regulated on day 30 of age, when comparing with that of day 0. Among these keratins, keratin 6 regulates collective keratinocyte migration (Wang et al., 2018). Keratin 7 is a type II intermediate filament protein which is primarily expressed in simple-type epithelia of glandular and ductal cells, as well as in stratified and transitional epithelia of urothelial and mesothelial tissues (Alam et al., 2021). Keratin 85 and keratin 75 are expressed primarily in hair follicles, nail beds, and lingual papillae (Yang et al., 2019). Keratin 80 facilitates the viability, proliferation, and migration of gastric cancer cells and inhibits cell apoptosis (Zhang et al., 2022). Thus, the role of these keratin proteins in rumen epithelium remains to be explored.

BH4 functions as a cofactor to convert amino acids to precursors of dopamine and serotonin (Li et al., 2015). In this study, we found DEPs significantly enriched in BH4 biosynthetic process, and pterin-4-alpha-carbinolamine dehydratase (PCD) and 4a-hydroxytetrahydrobiopterin dehydratase (DH) up-regulated on day 0 of age. PCD is required for enzymatic regeneration of BH4 (Citron et al., 1992). While, DH is the

rate-limiting enzyme for the recycling of BH4, and its activity significantly decreased in differentiated keratinocytes (Schallreuter et al., 1995). Thus, our data indicated that rumen epithelium of day 30 might contain more differentiated keratinocytes, while BH4 presents an important role in rumen epithelial keratinization.

The generation of BH4 is mediated by nicotinamide adenine dinucleotide (NADH) (Werner et al., 2011). In mitochondrion, BH4 is critical for maintaining its proper morphology (Zhou J. et al., 2019). Mitochondrion is a complex organelle that is essential in energy transduction and cellular signaling events. In rumen epithelium, mitochondria are majorly located in stratum basale cell (Graham and Simmons, 2005), which is considered to play a crucial role in the metabolism of SCFAs (Leighton et al., 1983). While the integrality of mitochondria morphology is also essential for cell differentiation, since it is usually accompanied by the activation of mitochondrial respiration, in which mitochondria exhibit a shift from glycolysis to oxidative phosphorylation (OXPHOS) (Shyh-Chang et al., 2013). The proton gradient across the cristae of mitochondria is the basis of OXPHOS, which provides the energy for cellular differentiation and proliferation (Ikon and Ryan, 2017). In this study, we found DEPs between rumen epithelium of day 0 and day 30 significantly enriched in molecular function of NADP binding, biological process of BH4 biosynthetic process and cristae formation. The proteins of mitochondrial part significantly up-regulated in day 30 of age, when the expression of mitochondrial contact site and cristae organizing system (MICOS) complex subunit and MICOS complex subunit MIC27 increased on day 30 of age. These data indicated an important role of mitochondrion during early calves rumen development, in which the biological processes of BH4 biosynthesis and molecular function of NADP binding participate in mitochondrial cristae formation.

Rumen epithelium, as the outermost barrier of calves, is responsible for preventing the invading of microbiota. Hence, the establishment of immune system in rumen epithelium is critical for the health of calves. In our study, DEPs between rumen epithelium of day 30 and day 60 significantly were enriched in NOD-like receptor signaling pathway, DEPs between rumen epithelium of day 60 and day 90 were significantly enriched in immune response, antigen processing and presentation and T-helper (Th) cell differentiation. NOD-like receptor signaling pathway can specifically sense and respond pathogen-associated molecular patterns and then secrete various cytokines by activating other signaling regulators (Liu et al., 2019). It is originally known as a critical regulator in immune responses, which specifically recognize pathogen-associated molecular patterns (Wilmanski et al., 2008). The immune system defends against pathogens and maintains the homeostasis of tissue for organism (Ganeshan and Chawla, 2014). It is suggested that CD4 T cells need to differentiate towards a Th1, Th2, Th17 to orchestrate a variety of adaptive immune responses (Butcher and Zhu, 2021). Th1 cells are primarily important in helping to mount a host defense against intracellular pathogens, whereas Th2 cells predominate in response to helminth infections and venoms (Romagnani, 1999). In addition, Th17 cells have been implicated not only in defense against opportunistic fungal or bacterial pathogens, but also in the pathogenesis of most common autoimmune diseases (Yang et al., 2014). These evidence indicated an important role of innate immune system during rumen epithelium development.

Rumen epithelium also plays roles in transporting urea, an end product of nitrogen metabolism. Urea, produced in the liver, can enter all major gastrointestinal tract compartments directly across the gastrointestinal tract wall (Hansen et al., 1976). In ruminant animals, it is transferred into the reticulo-rumen through diffusion across the rumen epithelium, where microbiota degrade recycled urea from saliva or rumen epithelium and then produce ammonia (Patra and Aschenbach, 2018). In calves, urea transporter driven urea recycling is one of the major mechanisms affecting rumen nitrogen metabolism (Stewart et al., 2005). Thus, establish the function of urea transport is of critical importance for rumen nitrogen recycle (Walpole et al., 2015). In this study, we found DEPs between day 30 and day 60 of age significantly enriched in Nitrogen metabolism and urea transmembrane transporter activity. The expression of urea transporter significantly increased on day 60 of age. These results indicated that the function of urea transport was well established after day 60 of age.

To conclude, our data indicated that keratinocyte differentiation, mitochondrion formation, the establishment of urea transport and innate immune system play central roles during rumen epithelium development. BH4 presents an important role in rumen epithelial keratinization. The biological processes of BH4 biosynthesis and molecular function of NADP binding participate in mitochondrial cristae formation. The proposed datasets provide a useful basis for future studies to better comprehend dairy calves rumen epithelial 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 below: Data are available via ProteomeXchange with identifier PXD037396.

Ethics statement

The animal study was reviewed and approved by Institutional Animal Care and Use Committees at the

Zhejiang Academy of Agricultural Sciences. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

XH and JJ initiated and directed the research; KZ designed and performed experiments, analyzed data and wrote the manuscript; YJ directed the research and was involved in sample collection, histological analysis and manuscript edition; SU, YC, and JW were involved in the animal experimentation and sample collection. All authors read and approved the final manuscript.

Funding

This work was supported by Department of Science and Technology of Zhejiang Province, China (Project No. LQ20C170002).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1071873/full#supplementary-material

References

Alam, C. M., Baghestani, S., Pajari, A., Omary, M. B., and Toivola, D. M. (2021). Keratin 7 is a constituent of the keratin network in mouse pancreatic islets and is upregulated in experimental diabetes. *Int. J. Mol. Sci.* 22 (15), 7784. doi:10.3390/ijms22157784

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., and Rasool, M. H. (2017). Proteomics: Technologies and their applications. *J. Chromatogr. Sci.* 55 (2), 182–196. doi:10.1093/chromsci/bmw167

Baldwin, R. L. T., and Connor, E. E. (2017). Rumen function and development. *Vet. Clin. North Am. Food Anim. Pract.* 33 (3), 427–439. doi:10.1016/j.cvfa.2017.06.001

Baldwin, R. L. t. (1998). Use of isolated ruminal epithelial cells in the study of rumen metabolism. *J. Nutr.* 128 (2), 293S–296S. doi:10.1093/jn/128.2.293S

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70 (2), 567–590. doi:10.1152/physrev.1990.70.2.567

Bragulla, H. H., and Homberger, D. G. (2009). Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *J. Anat.* 214 (4), 516–559. doi:10.1111/j.1469-7580.2009.01066.x

Butcher, M. J., and Zhu, J. (2021). Recent advances in understanding the Th1/Th2 effector choice. *Fac. Rev.* 10, 30. doi:10.12703/r/10-30

Citron, B. A., Davis, M. D., Milstien, S., Gutierrez, J., Mendel, D. B., Crabtree, G. R., et al. (1992). Identity of 4a-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transregulator of homeodomain proteins. *Proc. Natl. Acad. Sci. U. S. A.* 89 (24), 11891–11894. doi:10.1073/pnas.89.24.11891

Coulombe, P. A. (2017). The molecular revolution in cutaneous biology: Keratin genes and their associated disease: Diversity, opportunities, and challenges. *J. Invest. Dermatol* 137 (5), e67–e71. doi:10.1016/j.jid.2016.04.039

Diao, Q., Zhang, R., and Fu, T. (2019). Review of strategies to promote rumen development in calves. *Anim. (Basel)* 9 (8), 490. doi:10.3390/ani9080490

Finnegan, A., Cho, R. J., Luu, A., Harirchian, P., Lee, J., Cheng, J. B., et al. (2019). Single-cell transcriptomics reveals spatial and temporal turnover of keratinocyte differentiation regulators. *Front. Genet.* 10, 775. doi:10.3389/fgene.2019.00775

Ganeshan, K., and Chawla, A. (2014). Metabolic regulation of immune responses. *Annu. Rev. Immunol.* 32, 609–634. doi:10.1146/annurev-immunol-032713-120236

Graham, C., and Simmons, N. L. (2005). Functional organization of the bovine rumen epithelium. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 288 (1), R173–R181. doi:10.1152/ajpregu.00425.2004

Hansen, H. S., Rygard, J., and Engelholm, S. A. (1976). Clinical use of combined bleomycin and radiation therapy for head and neck tumours and testicular cancers. *Bull. Cancer* 63 (3), 371–378.

Ikon, N., and Ryan, R. O. (2017). Cardiolipin and mitochondrial cristae organization. *Biochim. Biophys. Acta Biomembr.* 1859 (6), 1156–1163. doi:10.1016/j.bbamem.2017.03.013

Ji, X., Tong, H., Settlage, R., Yao, W., and Jiang, H. (2021). Establishment of a bovine rumen epithelial cell line. *J. Anim. Sci.* 99 (10), skab273. doi:10.1093/jas/skab273

Leighton, B., Nicholas, A. R., and Pogson, C. I. (1983). The pathway of ketogenesis in rumen epithelium of the sheep. *Biochem. J.* 216 (3), 769–772. doi:10.1042/bj2160769

Li, W., Gong, M., Shu, R., Li, X., Gao, J., and Meng, Y. (2015). Molecular and enzymatic characterization of two enzymes BmPCD and BmDHPR involving in the regeneration pathway of tetrahydrobiopterin from the silkworm *Bombyx mori*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 186, 20–27. doi:10.1016/j.cbpb.2015.04.006

Liu, P., Lu, Z., Liu, L., Li, R., Liang, Z., Shen, M., et al. (2019). NOD-like receptor signaling in inflammation-associated cancers: From functions to targeted therapies. *Phytomedicine* 64, 152925. doi:10.1016/j.phymed.2019.152925

Meale, S. J., Chaucheyras-Durand, F., Berends, H., Guan, L. L., and Steele, M. A. (2017). From pre- to postweaning: Transformation of the young calf's gastrointestinal tract. *J. Dairy Sci.* 100 (7), 5984–5995. doi:10.3168/jds.2016-12474

Miller, W. F., Titgemeyer, E. C., Nagaraja, T. G., Watanabe, D. H. M., Felizari, L. D., Millen, D. D., et al. (2021). Influence of cane molasses inclusion to dairy cow diets during the transition period on rumen epithelial development. *Anim. (Basel)* 11 (5), 1230. doi:10.3390/ani11051230

Nishihara, K., Suzuki, Y., Kim, D., and Roh, S. (2019). Growth of rumen papillae in weaned calves is associated with lower expression of insulin-like growth factor-binding proteins 2, 3, and 6. *Anim. Sci. J.* 90 (9), 1287–1292. doi:10.1111/asj.13270

Patra, A. K., and Aschenbach, J. R. (2018). Ureases in the gastrointestinal tracts of ruminant and monogastric animals and their implication in urea-N/ammonia metabolism: A review. *J. Adv. Res.* 13, 39–50. doi:10.1016/j.jare.2018.02.005

Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: A hub for

mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50 (D1), D543–D552. doi:10.1093/nar/gkab1038

Roh, S. G., Suzuki, Y., Gotoh, T., Tatsumi, R., and Katoh, K. (2016). Physiological roles of adipokines, hepatokines, and myokines in ruminants. *Asian-Australas J. Anim. Sci.* 29 (1), 1–15. doi:10.5713/ajas.16.0001R

Romagnani, S. (1999). Th1/Th2 cells. *Inflamm. Bowel Dis.* 5 (4), 285–294. doi:10.1097/00054725-199911000-00009

Schallreuter, K. U., Lemke, K. R., Pittelkow, M. R., Wood, J. M., Korner, C., and Malik, R. (1995). Catecholamines in human keratinocyte differentiation. *J. Invest. Dermatol* 104 (6), 953–957. doi:10.1111/1523-1747.ep12606218

Shyh-Chang, N., Daley, G. Q., and Cantley, L. C. (2013). Stem cell metabolism in tissue development and aging. *Development* 140 (12), 2535–2547. doi:10.1242/dev.091777

Steele, M. A., Penner, G. B., Chaucheyras-Durand, F., and Guan, L. L. (2016). Development and physiology of the rumen and the lower gut: Targets for improving gut health. *J. Dairy Sci.* 99 (6), 4955–4966. doi:10.3168/jds.2015-10351

Stewart, G. S., Graham, C., Cattell, S., Smith, T. P., Simmons, N. L., and Smith, C. P. (2005). UT-B is expressed in bovine rumen: Potential role in ruminal urea transport. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 289 (2), R605–R612. doi:10.1152/ajpregu.00127.2005

Tozer, P. R., and Heinrichs, A. J. (2001). What affects the costs of raising replacement dairy heifers: A multiple-component analysis. *J. Dairy Sci.* 84 (8), 1836–1844. doi:10.3168/jds.S0022-0302(01)74623-1

Walpole, M. E., Schurmann, B. L., Gorka, P., Penner, G. B., Loewen, M. E., and Mutsvangwa, T. (2015). Serosal-to-mucosal urea flux across the isolated ruminal epithelium is mediated via urea transporter-B and aquaporins when Holstein calves are abruptly changed to a moderately fermentable diet. *J. Dairy Sci.* 98 (2), 1204–1213. doi:10.3168/jds.2014-8757

Wang, F., Chen, S., Liu, H. B., Parent, C. A., and Coulombe, P. A. (2018). Keratin 6 regulates collective keratinocyte migration by altering cell-cell and cell-matrix adhesion. *J. Cell Biol.* 217 (12), 4314–4330. doi:10.1083/jcb.201712130

Werner, E. R., Blau, N., and Thony, B. (2011). Tetrahydrobiopterin: Biochemistry and pathophysiology. *Biochem. J.* 438 (3), 397–414. doi:10.1042/BJ20110293

Wilmanski, J. M., Petnicki-Ocwieja, T., and Kobayashi, K. S. (2008). NLR proteins: Integral members of innate immunity and mediators of inflammatory diseases. *J. Leukoc. Biol.* 83 (1), 13–30. doi:10.1189/jlb.0607402

Xu, H. J., Zhang, Q. Y., Wang, L. H., Zhang, C. R., Li, Y., and Zhang, Y. G. (2022). Growth performance, digestibility, blood metabolites, ruminal fermentation, and bacterial communities in response to the inclusion of gallic acid in the starter feed of preweaning dairy calves. *J. Dairy Sci.* 105 (4), 3078–3089. doi:10.3168/jds.2021-20838

Xu, Q., Qiao, Q., Gao, Y., Hou, J., Hu, M., Du, Y., et al. (2021). Gut microbiota and their role in health and metabolic disease of dairy cow. *Front. Nutr.* 8, 701511. doi:10.3389/fnut.2021.701511

Yang, J., Sundrud, M. S., Skepner, J., and Yamagata, T. (2014). Targeting Th17 cells in autoimmune diseases. *Trends Pharmacol. Sci.* 35 (10), 493–500. doi:10.1016/j.tips.2014.07.006

Yang, X., Yamazaki, H., Yamakoshi, Y., Duverger, O., Morasso, M. I., and Beniash, E. (2019). Trafficking and secretion of keratin 75 by ameloblasts *in vivo*. *J. Biol. Chem.* 294 (48), 18475–18487. doi:10.1074/jbc.RA119.010037

Yuan, Y., Sun, D. M., Qin, T., Mao, S. Y., Zhu, W. Y., Yin, Y. Y., et al. (2022). Single-cell transcriptomic landscape of the sheep rumen provides insights into physiological programming development and adaptation of digestive strategies. *Zool. Res.* 43 (4), 634–647. doi:10.24272/j.issn.2095-8137.2022.086

Zhang, F., Wang, G., Yan, W., and Jiang, H. (2022). MiR-4268 suppresses gastric cancer Genesis through inhibiting keratin 80. *Cell Cycle* 21 (19), 2051–2064. doi:10.1080/15384101.2022.2085351

Zhang, X., Yin, M., and Zhang, L. J. (2019). Keratin 6, 16 and 17-critical barrier alarmin molecules in skin wounds and psoriasis. *Cells* 8 (8), 807. doi:10.3390/cells8080807

Zheng, K., Wu, J., Guo, L., Ying, Y., Li, P., Cao, Y., et al. (2021). The involvement of translationally controlled tumor protein during lamb rumen epithelium development. *Acta histochem.* 123 (5), 151737. doi:10.1016/j.acthis.2021.151737

Zheng, K., Yin, Y., Cao, Y., Guo, L., Li, P., Jiang, J., et al. (2022). Proteomic and parallel reaction monitoring approaches to evaluate biomarkers of mutton tenderness. *Food Chem.* 397, 133746. doi:10.1016/j.foodchem.2022.133746

Zhou, C. Y., Wang, C., Tang, C. B., Dai, C., Bai, Y., Yu, X. B., et al. (2019a). Label-free proteomics reveals the mechanism of bitterness and adhesiveness in Jinhua ham. *Food Chem.* 297, 125012. doi:10.1016/j.foodchem.2019.125012

Zhou, J., Xu, L., Duan, X., Liu, W., Zhao, X., Wang, X., et al. (2019b). Large-scale RNAi screen identified Dhpr as a regulator of mitochondrial morphology and tissue homeostasis. *Sci. Adv.* 5 (9), eaax0365. doi:10.1126/sciadv.aax0365

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1059913

OPEN ACCESS

EDITED BY
Shenghe Li,
Anhui Science and Technology University, China

REVIEWED BY
Hao Bai,
Yangzhou University, China
Mingxing Lei,
Chongqing University, China

*CORRESPONDENCE
Xuemei Deng,
deng@cau.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 02 October 2022
ACCEPTED 08 December 2022
PUBLISHED 04 January 2023

CITATION
Wang J, Cui K, Hua G, Han D, Yang Z, Li T, Yang X, Zhang Y, Cai G, Deng X and Deng X (2023), Skin-specific transgenic overexpression of ovine β-catenin in mice.
Front. Genet. 13:1059913.
doi: 10.3389/fgene.2022.1059913

Skin-specific transgenic overexpression of ovine β-catenin in mice

Jiankui Wang[1†], Kai Cui[1,2†], Guoying Hua[1], Deping Han[1], Zu Yang[1], Tun Li[1], Xue Yang[1], Yuanyuan Zhang[1], Ganxian Cai[1], Xiaotian Deng[1] and Xuemei Deng[1*]

[1]Key laboratory of Animal Genetics, Breeding and Reproduction of the Ministry of Agriculture, Beijing Key Laboratory of Animal Genetic Improvement, China Agricultural University, Beijing, China, [2]Key Laboratory of Feed Biotechnology, Ministry of Agriculture/Feed Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China

β-catenin is a conserved molecule that plays an important role in hair follicle development. In this study, we generated skin-specific overexpression of ovine β-catenin in transgenic mice by pronuclear microinjection. Results of polymerase chain reaction (PCR) testing and Southern blot showed that the ovine β-catenin gene was successfully transferred into mice, and the exogenous β-catenin gene was passed down from the first to sixth generations. Furthermore, real-time fluorescent quantitative PCR (qRT-PCR) and western blot analysis showed that β-catenin mRNA was specifically expressed in the skin of transgenic mice. The analysis of F6 phenotypes showed that overexpression of β-catenin could increase hair follicle density by prematurely promoting the catagen-to-anagen transition. The results showed that ovine β-catenin could also promote hair follicle development in mice. We, therefore, demonstrate domestication traits in animals.

KEYWORDS

β-catenin, ovine, transgenic mice, overexpression, hair follicles

Introduction

The skin is the largest organ in the body and provides a barrier to harmful substances (Proksch et al., 2008). Mammalian skin is composed of multiple layers of epithelial cells covered by a large number of hair follicles and glands (Watt, 2001). The dermal papilla is a relatively large structure in the basal part of a hair follicle (Watt, 2001). Once the dermal papilla is formed, the hair follicle structure becomes very stable throughout its life. This stable state, however, is based on a regular hair follicle cycle of growth (anagen), followed by retrogression (catagen) and then quiescence (telogen) (Li and Tumbar, 2021). Hair follicle formation during the embryonic stage depends on a large number of molecular signals. The interaction of these signaling molecules between the dermis and epithelial cells creates a complex regulatory network during hair follicle development (Hardy, 1992). In postnatal animals, the beginning of the hair follicle cycle is often accompanied by the initiation of a molecular signal in the hair papilla, which causes the transient proliferation

of bulge stem cells, further moving them to the root of the hair follicle (Oshima et al., 2001). Previous studies have identified many key molecules that regulate hair follicle development, such as fibroblast growth factors (FGFs), bone morphogenetic proteins (BMPs), and transforming growth factor β 2 (*TGF-β2*) (Millar, 2002). In addition, the Wnt signaling pathway is key in regulating hair follicle morphogenesis. *β-catenin,* the core molecule of the Wnt signaling pathway, plays an important role in hair follicle embryogenesis and *postpartum* development (Huelsken et al., 2001).

β-Catenin has two main functions. It activates the expression of target genes downstream of the Wnt pathway by activating the pathway and also mediates cell-cell adhesion (Schuman and Murase, 2003). *β-catenin* plays a diverse role in the Wnt signaling pathway and interacts with many proteins, such as *APC*, *AXIN*, and *GSK-3β* (Ozawa et al., 1989). In the classical Wnt signaling pathway, *β-catenin* is transported to the nucleus and binds to the *TCF/LEF* family, increasing the transcriptional activity (Doumpas et al., 2019). Previous studies have shown that *β-catenin* is essential for hair follicle development (Narhi et al., 2008). *β-catenin* mRNA expression increased during hair follicle morphogenesis at the embryonic stage (Tsai et al., 2014). In contrast, blocking *β-catenin* expression failed to form normal hair follicles (Liu et al., 2014). *In vitro* studies have indicated that knockout of *β-catenin* produces two phenotypes: a lack of substrate during hair follicle embryogenesis and hair loss during the first hair follicle cycle (Huelsken et al., 2001). Compared to that in wild-type mice, overexpression of incomplete *β-catenin* in mouse skin tissue could lead to an increase in the number and size of hair follicles and promote hair follicles to enter the anagen phase from the quiescent phase (Doumpaset al., 2019b). The "bulge-activation hypothesis" considers that when a signal is transferred from the dermal papilla and activates the stem cells in the bulge of the hair follicle, proliferation of the hair follicle stem cells is stimulated, and the follicles enter a new growth period. *β-catenin* plays a crucial role in this process (Li and Tumbar, 2021).

Aohan fine wool sheep, bred for their wool and meat, are well known for their slaughter rate and wool fineness (Liu et al., 2014). In recent years, the demand for superfine wool in the wool textile industry has increased rapidly. Improving wool density and fineness is a top priority in the production of high-quality wool (Han et al., 2021). In this study, we explored whether ovine *β-catenin* can be used as an exogenous gene to affect the development of hair follicles in mice, and the results from the transgenic mice model gave us a chance to understand the molecular mechanisms of skin-specific transgenic overexpression of ovine *β-catenin* in sheep (Wang et al., 2019). This will aid further research into the critical role of *β-catenin* in improving the wool quality in fine-wool sheep breeds.

Materials and methods

K14-β-catenin (ovine)-IRES-EGFP plasmid construction

The *β-catenin* was amplified by PCR from the cDNA of sheep skin, which contains SacII and BamHI sites, with the primer shown in Table 1. The purified products were inserted

TABLE 1 Primer sequences used to produce transgenic mice. F1, R1 for the amplification of β-catenin. F2, R2 for the amplification of promoter K14. F3, R3 for the qRT-PCR of transgenic mice. F4, R4 for the amplification of housekeeping gene. F5, R5 for the identification of positive mice. F6, R6 for the synthesis of probe of Southern blot. The underlined bases are the restrict enzyme sites.

Primers	Sequence (5′-3′)	Fragment length	Ta value(C)	
β-*catenin*	F1	TCCCCGCGGCGGAGACGGAGCAAGGT	2656bp	56
	R1	CGCGGATCCGCAAGCAAAGTCAGTACCAT		
K14	F2	CATTAATATCCCTGCAGAAGAAGGAGAC	2000bp	55
	R2	ATAAGCGCTGGCTGAGTGAAGAGAAGG		
qRT-PCR	F3	AGCGTCGTACATCTATGGG	368bp	59
	R3	ATAATCCTGTGGCTTGACC		
GAPDH	F4	GTCCGTTGTGGATCTGACCT	245bp	59
	R4	TGCTGTAGCCGAATTCATTG		
Identification	F5	TCACCGTAGGGAGGAAAT	1226bp	56
	R5	TAGCGTCTCAGGGAACAT		
Probe synthesis	F6	CAAGAAAGCCCAAAACAC	739bp	57
	R6	TAGCGTCTCAGGGAACAT		

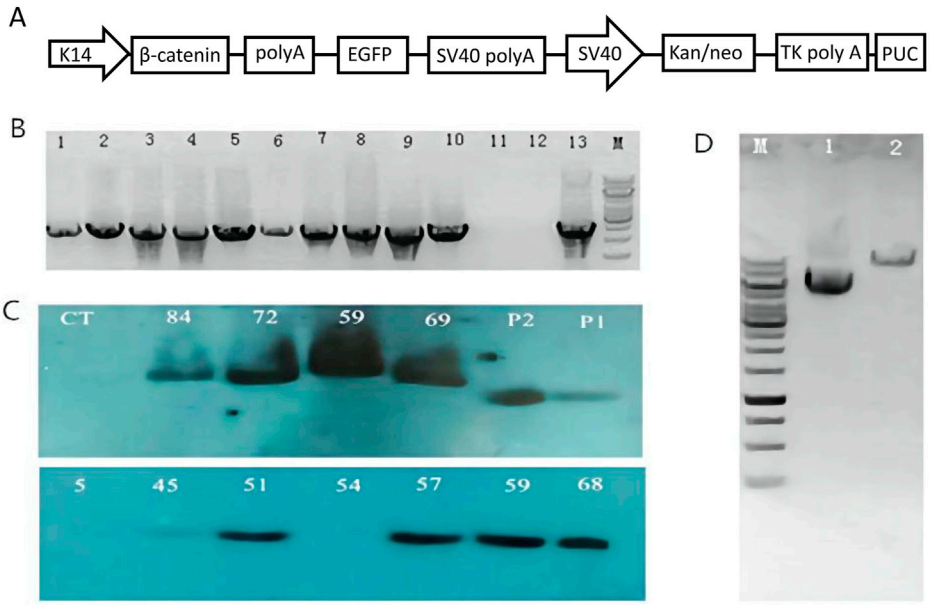


FIGURE 1
Generation of ovine β-catenin transgenic mice (F0). **(A)** Schematic diagram of transgenic vector (pK14-β-catenin-EGFP). **(B)** Agarose gel analysis on PCR products from genomic DNA extracted by transgenic founder mice tail. The PCR products from mouse number 5, 45, 51, 54, 57, 59, 68, 69, 72, and 84 are shown in well #1 to #10 respectively; the PCR product with H_2O as a control temple is shown in well #11; the negative PCR product with DNA extracted from wild type mouse tail is in well #12; PCR product using transgenic vector DNA as template is shown in well#13. **(C)** Southern blot with DNA from transgenic mice skin. WT, a sex- and age-matched wild-type control; p1, p2 plasmid controls. **(D)** Analysis of β-catenin Vector DNA by AseI digestion on 1% agarose gel. Lane 1: Vector DNA without any restriction enzyme digestion; Lane 2: Vector DNA digested by AseI; M: 1 Kb DNA ladder.

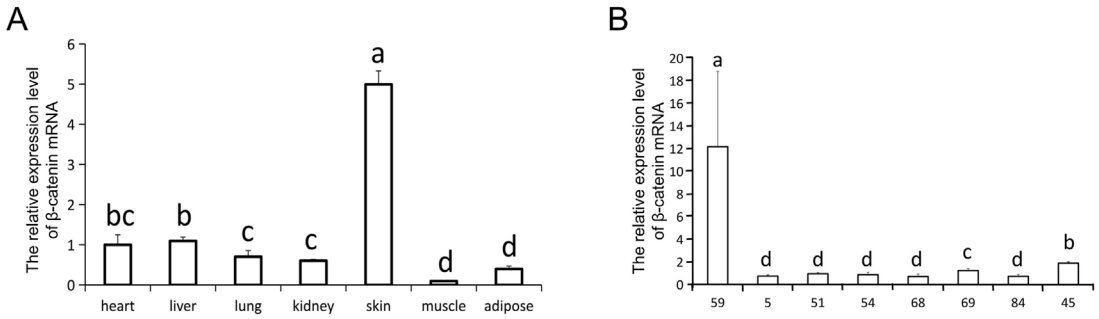


FIGURE 2
The relative expression level of β-catenin mRNA. **(A)** The expression of β-catenin mRNA in skin of different transgenic mouse lines (F0). 45 is a wilde-type control line. **(B)** qRT-PCR analysis of the β-catenin mRNA expression in different tissues of line 59 (F0). The relative quantity of β-catenin mRNA in **(A)** and **(B)** were determined *via* qRT-PCR using $2^{-\Delta\Delta CT}$ method, with GAPDH as the internal control. Each experimental group contained three replicates and qRT-PCR was performed in triplicate for each sample. Bars with common lowercase letters had no significant difference at the 5% level.

into PMD19T plasmid vector (TAKARA, Japan) and sequenced at SinoGenoMas Co., Ltd. (Beijing, China). The 2.0 kb K14 promoter with AseI and AfeI sites obtained by PCR was cloned into PMD-19T vector and sequenced as above. In order to remove the CMV promoter of pIRES-EGFP, the restriction enzyme AfeI and AseI were used to digest and the human K14 promoter was inserted into the AseIand AseI site to produce the skin-specific expression plasmid K14-IRES-EGFP. Afterwards, the ovine *β-catenin* gene was inserted into K14-IRES-EGFP plasmid by digesting with

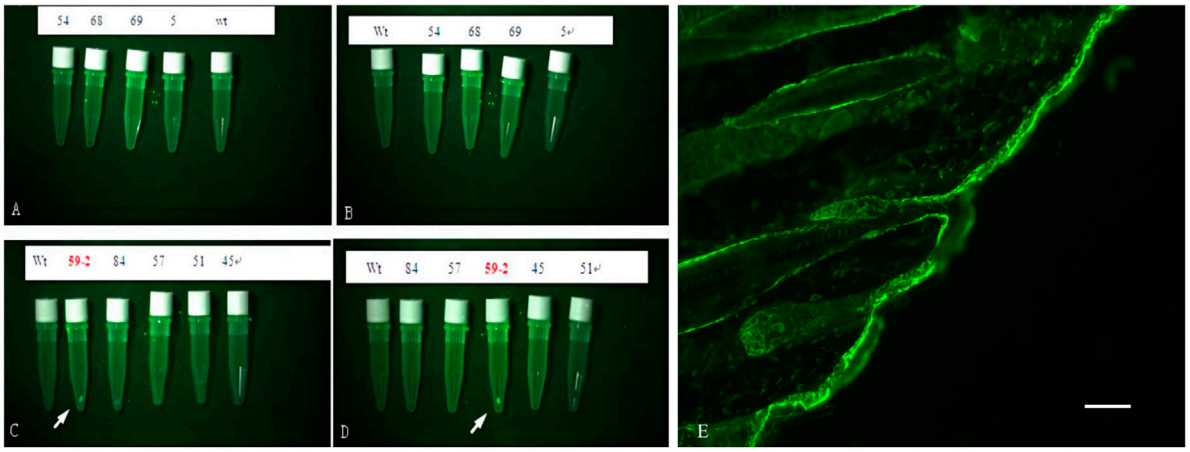


FIGURE 3
The detection of Green Fluorescent Protein (GFP) in skin of F0 transgenic mice. **(A)**, **(B)**, **(C)** and **(D)** are the fluorescence microscope analysis of the skin of the transgenic mice. Wt is wilde-type control. GFP was detected in line 59, indicated with white arrow. **(E)** The fluorescence Protein in skin tissue of ovine β-catenin transgenic mice, the scale bar: 60 μm.

restriction enzyme BamHI and SacII. The resultant plasmid K14-*β-catenin*-IRES-EGFP was then sequenced (SinoGenoMax Co., Ltd., Beijing, China).

Generation of transgenic mice

In order to generate the linear DNA fragments, the AseI was used to digest the plasmid K14-*β-catenin*-IRES-EGFP, then the restricted DNA products were resuspended in sterile ddH_2O at a concentration of 20 ng/μL. Then, the products were injected into the male pronuclei of zygote of C57BL/6 by Shanghai Research Center for Model Organisms. The F1 generation mouse were produced by mating positive founder mouse (carry K14-β-catenin-IRES-EGFP fragments) with wild type.

Polymerase chain reaction and southern blot analysis of F0 and F1 mice

Genomic DNA was extracted from tail of the transgenic/control mouse by using the TIANamp Genomic DNA Kit (TIANGEN, Beijing, China). PCR assay was performed by using the primers shown in Table 1. The 1226 bp PCR product contains partial K14 promoter and partial ovine *β-catenin* DNA sequence. The PCR procedure: 5 min at 95°C, 35 cycles of 30 s for 95°C, 30 s for 58°C and 72°C for 1 min, 72°C for 10 min and hold at 4°C forever. The further southern blot assay, digestion: the BstEII was selected for digested the DNA (60°C, 20 h), which is obtained from mouse tail. The probe (739 bp) amplified by primer (F6 and R6 in Table 1) across promoter and β-catenin sequence, was labeled with the PCR digoxigenin probe synthesis kit (Roche). The DIG-High Prime DNA Labeling and Detection Starter KitII was used for washing and hybridization (Roche, Basel, Switzerland).

Analysis of transgene expression of F2 and F3 mice

Total RNA of different tissue (heart, liver, lung, kidney, skin, muscle and adipose) were isolated by using an RNAprep Pure Tissue Kit (TIANGEN, Beijing, China). According to the guidelines in the kit, 1–2 μg total RNA was used to synthesized the first-stand cDNA using ImPromII Reverse Transcription System (Promega, Beijing, China). qRT-PCR was performed (primers in Table 1) and the analysis of *β-catenin* expression was evaluated by quantitative real-time PCR using real-time PCR detection system (Bio-Rad, United States). The *β-actin* was used as internal references. The results of *β-catenin* expression related *β-actin* were test in triplicate. The final results were generated using the $2^{-\triangle\triangle CT}$ method (Livak and Schmittgen, 2001).

Protein analysis of F3 mice

Samples from the tail of transgenic mice were used for fluorescence microscope analysis of green fluorescent protein (GFP). For western blot analysis, according to the kit guidelines of RIPA buffer (BiYoTimel, Suzhou, China), protein was extracted from different tissues of transgenic and control mouse. The concentration of extracted protein was measured by BCA Protein Assay Kit (BiYoTime, China) and the bovine

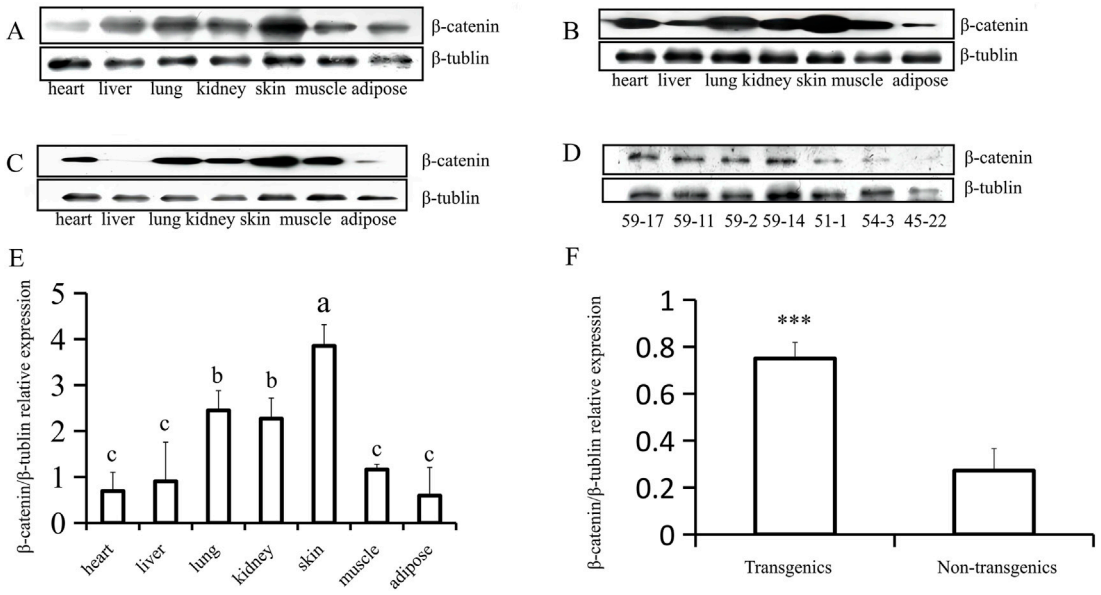


FIGURE 4
Western-blot analysis of β-catenin protein expression in heart, liver, lung, kidney, skin, muscle and adipose (F1). **(A) (B)** and **(C)** The analysis of β-catenin protein expression in different tissues of three transgenic mice (59−17, 59−2, and 59−11). **(D)** The expression of β-catenin protein in transgenic (59−17, 59−11, and 59−2) and non-transgenic mice (51−1, 54−3, and 45−22). **(E)** The statistics of β-catenin protein expression in different tissues of transgenic mouse lines, the bars with common lowercase letters had no significant difference at the 5% level. **(F)** Comparative analysis of β-catenin protein expression between transgenic and non-transgenic mice. ***, $p < 0.001$.

serum albumin was used as a standard. Each lane contained an equal amount of protein (20 μg per lane) and was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Then transferred onto the polyvinyl-difluoride membranes at 120 V for 90min. The membranes were incubated with 1:1000 primary anti-β-catenin (abcam, E247, United States), and β-actin (Beyotime, AF5001, China). After 1 hour blocking, wash five times with Tris-buffered saline, each time last 5 min. Horseradish peroxidase-conjugated secondary antibody (1; 10,000) was used to incubated the membranes (BiYoTime, Suzhou, China). Then, the membranes were visualized by using enhanced chemiluminescence system after wased six times with TBST (each time last 5 min).

Histology analysis of F6 mice

The 8 micron-thick of skin sections were obtained by freezing microtome (LEICA, CM 1900, Germany). The hematoxylin and eosin staining was performed using a Hematoxylin and Eosin Staining Kit (Beyotime, China, C0105S). For analysis of the hair follicle density, the Zeiss dissecting microscope and Zesiss light microscope (Carl; Zeiss, Germany) were used, the hair follicle numbers of per one square millimeter of mice back skin section was counted and the average of three areas of each samples were identified as the hair follicle density.

Statistical analysis

Differences between transgenic and non-transgenic mice groups were tested using independence Student's t-test. In this study, two-sided p values were identified significant while lower than 0.05.

Results

Identification of F0 and F1 transgenic mice

In this study, a pK14-β-catenin-IRES-GFP vector was successfully constructed (Figure 1A). The vector was linearized by Ase I digestion and then microinjected into the nucleus of mouse zygotes (Figure 1D). 480 fertilized oocytes were injected with linearized vector fragments, and the injected oocytes were then transplanted into 24 recipient mice. A total of 93 founder mice were obtained, of which 10 were identified as transgenic mice by PCR (Figure 1B). The presence of the exogenous ovine *β-catenin* gene was further confirmed by Southern blot (Figure 1C), and eight of the ten PCR-positive mice were confirmed to be transgenic mice carrying the ovine *β-catenin* gene. Eight transgenic mice were healthy and defect-free. Eight transgenic lines were established by crossing the founders with wild-type C57BL/

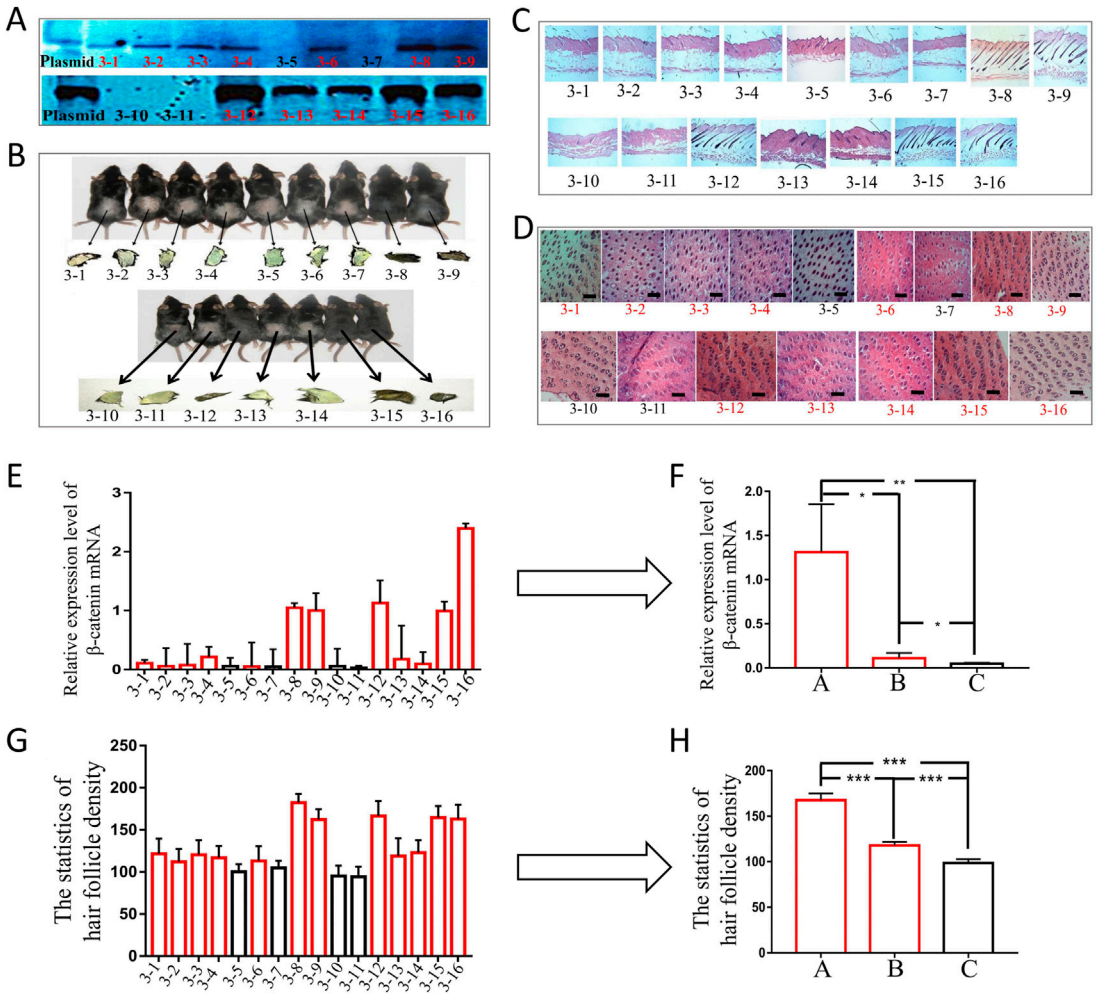


FIGURE 5
The phenotypic analysis of F6 transgenic mice by overexpressing ovine β-catenin. **(A)** Southern-blot of two transgenic families, the thicker black bands represent higher copy number of foreign gene (ovine β-catenin) while the thinner bands represent lower copy number of ovine β-catenin. **(B)** The two transgenic families of mice and their skin color. **(C)** The longitudinal slices of transgenic mice skin of the two families. **(D)** The transection sections of transgenic mice skin of the two families, the scale bar: 600 μm. **(E)** The expression level of β-catenin mRNA in transgenic mice of the two families. **(F)** The analysis of β-catenin mRNA levels of transgenic mice that were divided into three groups (A: high copy number group, **(B)** low copy number group and **(C)** non-transgenic siblings). *, $p < 0.05$; **, $p < 0.01$. **(G)** Statistical analysis of hair follicle density in transgenic mice of the two families. **(H)** The analysis of hair follicle density of transgenic mice that were divided into three groups (A: high copy number group, **(B)** low copy number group and **(C)** non-transgenic siblings). ***, $p < 0.001$.

6 mice. It was then expanded to the sixth generation. In this process, PCR and Southern blot techniques were used to identify transgenic mice.

Characterization of ovine *β-catenin* expression in transgenic mice

To assess the characteristics of ovine *β-catenin* expression in transgenic mice, we used qRT-PCR to detect *β-catenin* mRNA levels in different tissues of transgenic mice. The results showed that the expression level of *β-catenin* mRNA in the skin of transgenic mice of line 59 was significantly higher than that in the skin of transgenic mice of other lines ($p < 0.01$) (Figure 2A). We investigated the tissue specificity of ovine *β-catenin* transgene expression in F3 transgenic mice of line 59, and the results showed the transgene expression level in the skin was significantly higher than in other tissues ($p < 0.01$) (Figure 2B). The results indicated that the transgenic mouse of F1 could pass the foreign ovine *β-catenin* onto their offspring, and ovine *β-catenin* was specifically expressed in the skin of transgenic mice.

Tissue-specific expression of the *β-catenin* protein in transgenic mice

Analysis using fluorescence microscopy showed that green fluorescent protein (GFP) was expressed in the skin tissue in mice from line 59 but not in the other lines (Figures 3A–E). The western blot results show that the *β-catenin* protein could be detected in the heart, liver, lung, kidney, skin, muscle, and adipose tissues (Figures 4A–C). The highest expression level of the *β-catenin* protein was observed in the skin compared with that in other tissues ($p < 0.01$). These results indicated that the *β-catenin* protein expression level in the skin of transgenic mice of line 59 was significantly higher than that in the other lines (Figures 4D,F) ($p < 0.01$). In addition, the *β-catenin* protein was expressed at significantly higher levels in the skin of transgenic mice than in that of non-transgenic mice (Figures 4A–C,E).

The significant dosage effect of *β-catenin* copy number and hair follicle traits

To carry out phenotypic analysis, we selected blot transgenic and non-transgenic mice of the same age and sex from three families of F6. Southern was used to not only identify the transgenic mice but also determine the different copy numbers among transgenic mice by the depth of band gray levels (Figure 5A). The results indicated that the greater the copy number of the foreign gene (ovine *β-catenin*) between transgenic and non-transgenic mice, the more significant the differences in *β-catenin* expression (Figures 5E,F) and hair follicle density (Figures 5D,G,H). In addition, transgenic mice with a higher copy number of ovine *β-catenin* showed an earlier anagen phase on postnatal day 56 (telogen) of the dorsal skin (Figures 5B,C).

Discussion

In our previous study, we cloned the *β-catenin* cDNA sequence (GenBank accession number KC668410) from Aohan fine wool sheep. Homology analysis showed that the *β-catenin* gene is highly conserved among species, and the amino acid sequence of the *β-catenin* protein is 99% conserved among species (Cui et al., 2014). To further explore whether ovine *β-catenin* can be used as a foreign gene to affect hair follicle development, we generated transgenic mice and expressed the ovine *β-catenin* protein specifically in their skin tissue. The results of the qRT-PCR analysis showed that *β-catenin* mRNA could be detected in various tissues. This was because the primers cannot distinguish the endogenous and exogenous *β-catenin* sequences since sheep and mouse share 89.39% of the CDS sequence of *β-catenin*. The expression of *β-catenin* mRNA in other tissues may be a result of endogenous *β-catenin*. Of the different tissues tested, *β-catenin* mRNA was most prevalent in the skin, indicating that the foreign gene was successfully expressed in the target tissue. The western blot and qRT-PCR results suggested that the *β-catenin* protein can be passed down from F1 to their offspring and that *β-catenin* protein is specifically expressed in the skin of the transgenic mice. Green fluorescent protein detected in the skin of line 59 showed that the internal ribosome entry site (IRES) mediated the exogenous gene and GFP expression together, which could improve the detection accuracy of transgenic-positive animals (Kozak, 2005).

Hair follicles are formed during embryogenesis and show regular hair follicle cycles throughout their lifecycle (anagen, catagen, and telogen) (Hardy, 1992; Paus et al., 1999). As a key molecule in the Wnt signaling pathway, *β-catenin* is highly expressed in hair follicles (Ozawa et al., 1989; Goding, 2000). Transient activation of *β-catenin* signaling in skin keratinocytes can promote hair follicle growth (Van Mater et al., 2003). Frozen sections of skin showed that the overexpression of *β-catenin* in the skin can enhance follicle density, similar to previous research (Gat et al., 1998). Immunohistochemical analysis of transgenic mice showed that *β-catenin* was expressed in the dermal papilla and inner root sheath of the hair follicles (Wang et al., 2012). Previous research has revealed that *β-catenin* is highly conserved during biological evolution (Fuchs et al., 2001). The skin-specific expression of ovine *β-catenin* can increase wool follicle density, leading to higher wool production in transgenic sheep (Wang et al., 2019), which agrees with the results of this study.

In conclusion, we successfully generated transgenic mice with skin-specific overexpression of ovine *β-catenin*. The recombinant plasmid exhibited good genetic stability, and the exogenous gene could be transmitted from F1 to F6. The transgenic mice were healthy and had no defects. We demonstrated that transgenic mice overexpressing ovine *β-catenin* showed a high density of hair follicles and an early-entered anagen (Grigoryan et al., 2008).

Data availability statement

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

Ethics statement

The animal study was reviewed and approved by the Ethics of Animal Experiments of China Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

XD designed the research. JW and KC performed the experiments. JW, KC, and GH analyzed the data. XD, KC, and JW wrote and revised the manuscript. Other authors participated in manuscript revision or sample collection.

Funding

This work was supported by programs of National Key R&D Program of China (2021YFF1000700), the Major Project for Cultivation Technology of New Varieties of Genetically Modified Organisms of the Ministry of Agriculture (2016ZX08008-001), the National Natural Science Foundation of China (32002145).

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

Cui, K., Yang, Z., Darwish, H., Zhang, Y., Ge, Y., Zhang, X., et al. (2014). Molecular cloning and characterization of the beta-catenin gene from fine-wool sheep. *Gene* 546, 277–282. doi:10.1016/j.gene.2014.05.067

Doumpas, N., Lampart, F., Robinson, M. D., Lentini, A., Nestor, C. E., Cantu, C., et al. (2019). TCF/LEF dependent and independent transcriptional regulation of Wnt/β-catenin target genes. *Embo J.* 38, e98873. doi:10.15252/embj.201798873

Fuchs, E., Merrill, B. J., Jamora, C., and DasGupta, R. (2001). At the roots of a never-ending cycle. *Dev. Cell* 1, 13–25. doi:10.1016/s1534-5807(01)00022-3

Gat, U., DasGupta, R., Degenstein, L., and Fuchs, E. (1998). De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin. *Cell* 95, 605–614. doi:10.1016/s0092-8674(00)81631-1

Goding, C. R. (2000). Mitf from neural crest to melanoma: Signal transduction and transcription in the melanocyte lineage. *Genes Dev.* 14, 1712–1728. doi:10.1101/gad.14.14.1712

Grigoryan, T., Wend, P., Klaus, A., and Birchmeier, W. (2008). Deciphering the function of canonical Wnt signals in development and disease: Conditional loss- and gain-of-function mutations of beta-catenin in mice. *Genes Dev.* 22, 2308–2341. doi:10.1101/gad.1686208

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

Hardy, M. H. (1992). The secret life of the hair follicle. *Trends Genet.* 8, 55–61. doi:10.1016/0168-9525(92)90350-d

Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G., and Birchmeier, W. (2001). Beta-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* 105, 533–545. doi:10.1016/s0092-8674(01)00336-1

Kozak, M. (2005). A second look at cellular mRNA sequences said to function as internal ribosome entry sites. *Nucleic Acids Res.* 33, 6593–6602. doi:10.1093/nar/gki958

Li, K. N., and Tumbar, T. (2021). Hair follicle stem cells as a skin-organizing signaling center during adult homeostasis. *Embo J.* 40, e107135. doi:10.15252/embj.2020107135

Liu, N., Li, H., Liu, K., Yu, J., Bu, R., Cheng, M., et al. (2014). Identification of skin-expressed genes possibly associated with wool growth regulation of Aohan fine wool sheep. *BMC Genet.* 15, 144. doi:10.1186/s12863-014-0144-1

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

Millar, S. E. (2002). Molecular mechanisms regulating hair follicle development. *J. Invest. Dermatol* 118, 216–225. doi:10.1046/j.0022-202x.2001.01670.x

Narhi, K., Jarvinen, E., Birchmeier, W., Taketo, M. M., Mikkola, M. L., and Thesleff, I. (2008). Sustained epithelial beta-catenin activity induces precocious hair development but disrupts hair follicle down-growth and hair shaft formation. *Development* 135, 1019–1028. doi:10.1242/dev.016550

Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K., and Barrandon, Y. (2001). Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* 104, 233–245. doi:10.1016/s0092-8674(01)00208-2

Ozawa, M., Baribault, H., and Kemler, R. (1989). The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *Embo J.* 8, 1711–1717. doi:10.1002/j.1460-2075.1989.tb03563.x

Paus, R., Muller-Rover, S., Van Der Veen, C., Maurer, M., Eichmuller, S., LinG, G., et al. (1999). A comprehensive guide for the recognition and classification of distinct stages of hair follicle morphogenesis. *J. Invest. Dermatol* 113, 523–532. doi:10.1046/j.1523-1747.1999.00740.x

Proksch, E., Brandner, J. M., and Jensen, J. M. (2008). The skin: An indispensable barrier. *Exp. Dermatol* 17, 1063–1072. doi:10.1111/j.1600-0625.2008.00786.x

Schuman, E. M., and Murase, S. (2003). Cadherins and synaptic plasticity: Activity-dependent cyclin-dependent kinase 5 regulation of synaptic beta-catenin-cadherin interactions. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 358, 749–756. doi:10.1098/rstb.2002.1256

Tsai, S. Y., Sennett, R., Rezza, A., Clavel, C., Grisanti, L., Zemla, R., et al. (2014). Wnt/β-catenin signaling in dermal condensates is required for hair follicle formation. *Dev. Biol.* 385, 179–188. doi:10.1016/j.ydbio.2013.11.023

Van Mater, D., Kolligs, F. T., Dlugosz, A. A., and Fearon, E. R. (2003). Transient activation of beta -catenin signaling in cutaneous keratinocytes is sufficient to trigger the active growth phase of the hair cycle in mice. *Genes Dev.* 17, 1219–1224. doi:10.1101/gad.1076103

Wang, H. D., Yang, L., Yu, X. J., He, J. P., Fan, L. H., Dong, Y. J., et al. (2012). Immunolocalization of beta-catenin and Lef-1 during postnatal hair follicle development in mice. *Acta histochem.* 114, 773–778. doi:10.1016/j.acthis.2012.01.004

Wang, J., Cui, K., Yang, Z., Li, T., Hua, G., Han, D., et al. (2019). Transcriptome analysis of improved wool production in Skin-Specific transgenic sheep overexpressing ovine beta-Catenin. *Int. J. Mol. Sci.* 20, 620. doi:10.3390/ijms20030620

Watt, F. M. (2001). Stem cell fate and patterning in mammalian epidermis. *Curr. Opin. Genet. Dev.* 11, 410–417. doi:10.1016/s0959-437x(00)00211-2

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fvets.2023.1085474

OPEN ACCESS

EDITED BY
Sihua Jin,
Anhui Agricultural University, China

REVIEWED BY
Hui Diao,
Sichuan Animal Science Academy, China
Hailiang Song,
Beijing Academy of Agricultural and Forestry Sciences, China

*CORRESPONDENCE
Hong Hu
✉ haiyanghh@163.com

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

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

RECEIVED 31 October 2022
ACCEPTED 03 January 2023
PUBLISHED 03 February 2023

CITATION
Wen A, Zhang W, Xu J, Wang K and Hu H (2023) Identification and characterization of extrachromosomal circular DNA in Wei and Large White pigs by high-throughput sequencing. *Front. Vet. Sci.* 10:1085474. doi: 10.3389/fvets.2023.1085474

Identification and characterization of extrachromosomal circular DNA in Wei and Large White pigs by high-throughput sequencing

Aiyou Wen[1†], Wei Zhang[2†], Jingen Xu[1], Kunping Wang[1] and Hong Hu[1*]

[1]College of Animal Science, Anhui Science and Technology University, Chuzhou, China, [2]Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei, China

Wei pig (WP) and Large White pig (LP) are fatty and lean breeds, respectively. Extrachromosomal circular DNA (eccDNA) plays an important role in regulating signaling pathway processes of cell. However, there are few reports regarding the eccDNA and ecDNA profiles in WP and LP. The present work aimed to investigate the eccDNA and ecDNA profiles between WP and LP. Three WPs and three LPs (100 ± 1.3 kg) were selected for analysis of eccDNA and ecDNA in the ear samples. Results showed that there were 39,686,953,656–58,411,217,258 and 53,824,168,657–58,311,810,737 clean data for WP and LP, respectively. Sequencing yielded 15,587–25,479 and 71,123–79,605 eccDNAs from the ear samples of WP and LP, respectively. There were 15,111 and 22,594 eccDNA-derived genes in the WP and LP, respectively, and 13,807 eccDNA-derived genes were common in the ear samples of both pigs. Sequencing yielded 13–19 and 27–43 ecDNAs in the ears of WP and LP, respectively. There were 1,005 and 1,777 ecDNA-derived genes in WP and LP, respectively, and 351 ecDNA-derived genes were common in the ear samples of both pigs. The most significant KEGG pathways of eccDNA-derived genes were axon guidance, focal adhesion, metabolic pathways, MAPK signaling pathway, Hedgehog signaling pathway, microRNAs in cancer, tight junction, phospholipase D signaling pathway, endocytosis, and sphingolipid signaling pathway. Furthermore, the most significant KEGG pathways of ecDNA-derived genes were olfactory transduction, B cell receptor signaling pathway, and chemical carcinogenesis. The eccDNA00044301 was lower abundance, while the ecDNA00000060 was higher abundance in WP compared with that in LP. Summary, we found that eccDNAs and ecDNAs are common in WP and LP and occur in sizes large enough to carry one or several partial or complete genes. These findings have expanded the knowledge repertoire of circular DNA in pig and will provide a reference for the use of pigs as a medical model and help discovery of new genetic markers to select high-quality breeds.

KEYWORDS

eccDNA, ecDNA, Wei pig, Large White pig, high-throughput sequencing

Introduction

Extrachromosomal circular DNA (eccDNA), an unconventional presence of extrachromosomal DNA, is a circular DNA molecule derived from genomic DNA (1, 2). It can be categorized as microDNA, spcDNA, and ecDNA. The ecDNA which found in cancerous cell are double-stranded circular DNA molecule ranging from a few 100 kb to several Mbs in size (3, 4). It has recently been shown that eccDNA plays an important role in regulating biological processes including signal pathway, telomere trimming, and stress response (5, 6). The special distribution of eccDNA enhances the ability to use it as a biomarker for some diseases, especially cancer (7). With the development of high-throughput sequencing, eccDNA

sequencing may be helpful for researchers to further explore the potential biological functions of eccDNA.

The pig is an ancient omnivorous mammal. As an economically important animal in agriculture, pigs are a main source of meat products and are in high demand worldwide (8, 9). The similarities between pig and human, in terms of genome, immunology, and, anatomical structure, enhance its usefulness as a biological model (10–12). Pigs have become an important biomedical research model for studying obesity, diabetes, hypertension, and other human diseases (13).

The Wei pig (WP) is a famous fatty breed in Anhui Province of China; it is resistant to stress and has a high reproductive rate and fat content (14). The characteristics of Large White pig (LP, an excellent lean breed) include a fast growth speed, high lean rate, and low feed gain ratio (14, 15). There are some differences between the two varieties owing to differences in their genetic backgrounds. However, few reports are available regarding the differences in eccDNA and ecDNA between WP and LP. The present work used high-throughput sequencing to determine eccDNA and ecDNA profiles from the ear samples of WP and LP. Bioinformatics was used to assess the similarities and differences of eccDNA and ecDNA between WP and LP. This information could provide a reference for the potential use of pigs as a medical model and help discover new genetic markers to select high-quality breeds.

Materials and methods

Experiment design

This experiment was approved by the Animal Care and Use Committee of Anhui Science and Technology University. Three male WPs and three male LPs (weight 100 ± 1.3 kg) were selected from cooperative farm of Anhui Science and Technology University were dived into CW and CY groups. The ear samples were collected from these pigs and stored at −70°C until the DNA extraction procedure.

Library construction, sequencing, and circular DNA analysis

The DNA of ear sample was extracted use a commercial kit (Tiangen, China) and the quality was detected using Qubit (ThermoFisher, USA) and NanoDrop (ThermoFisher, USA). The library construction and sequencing were carried out according to the methods described by Møller et al. (16). The samples were processed by exonuclease V (NEB, USA) to degrade the linear genomic DNA For the restriction enzyme-based approach, circular DNA were digested using MspI (NEB, USA). Then, the library was sequenced on Novaseq 6000 (Illumina, USA) and 150 paired-end reads were generated (GeneDenovo, China). The high-quality clean reads were obtained from raw reads by Ffastp software. Q20 and Q30 were used to determine the quality of circular DNA. Large circular DNA in tumors is generally > 100 kb, while the other types of circular DNA are usually below 100 kb. Therefore, we used 100 kb as a threshold to distinguish eccDNA from ecDNA. The corresponding correlation analysis was performed independently.

Analysis of cDNA distribution and its derived genes

EccDNA and ecDNA were annotated according to its source region. We divided the genome into exons, introns, gene_up2k, gene_down2k, and intergenic regions and annotated each eccDNA and ecDNA based on the region to which it belonged. If the eccDNA and ecDNA spanned multiple regions, it was classified according to the following priority: exon > intron > gene_ up2k > gene_ down2k > intergenic. The eccDNA and ecDNA related genes were derived from coding gene regions. The common and unique eccDNA and ecDNA related gene between CW and CY groups were analyzed by Venn analysis (https://www.omicshare.com).

Differential abundance analysis of eccDNA and ecDNA

The differential abundance of eccDNA and ecDNA were analyzed using EdgeR software. The eccDNA and ecDNA with a p-value < 0.05 and a $|\log_2 FC| > 1$ was regard as the significant difference between CW and CY groups.

Analysis of GO and KEGG

The enriched pathway of eccDNA and ecDNA related genes were performed using GO and KEGG database in www.omicshare.com.

TABLE 1 Summary of high-throughput sequencing.

Pig	Raw data (bp)	Clean data (bp)	Q20 (%)	Q30 (%)	Total mapped (%)	GC (%)
CW1	72,057,443,400	58,301,151,414	97.50%	93.54%	80.51	48.22
CW2	67,214,298,000	58,411,217,258	97.24%	93.46%	79.79	50.90
CW3	47,459,279,100	39,686,953,656	97.62%	94.04%	81.08	50.23
CY1	58,233,148,200	54,898,548,653	96.81%	92.54%	79.66	52.40
CY2	58,510,024,200	53,824,168,657	96.42%	92.18%	79.91	56.80
CY3	61,656,094,500	58,311,810,737	96.86%	92.48%	84.75	52.79

CW, Wei pig; CY, Large White pig.

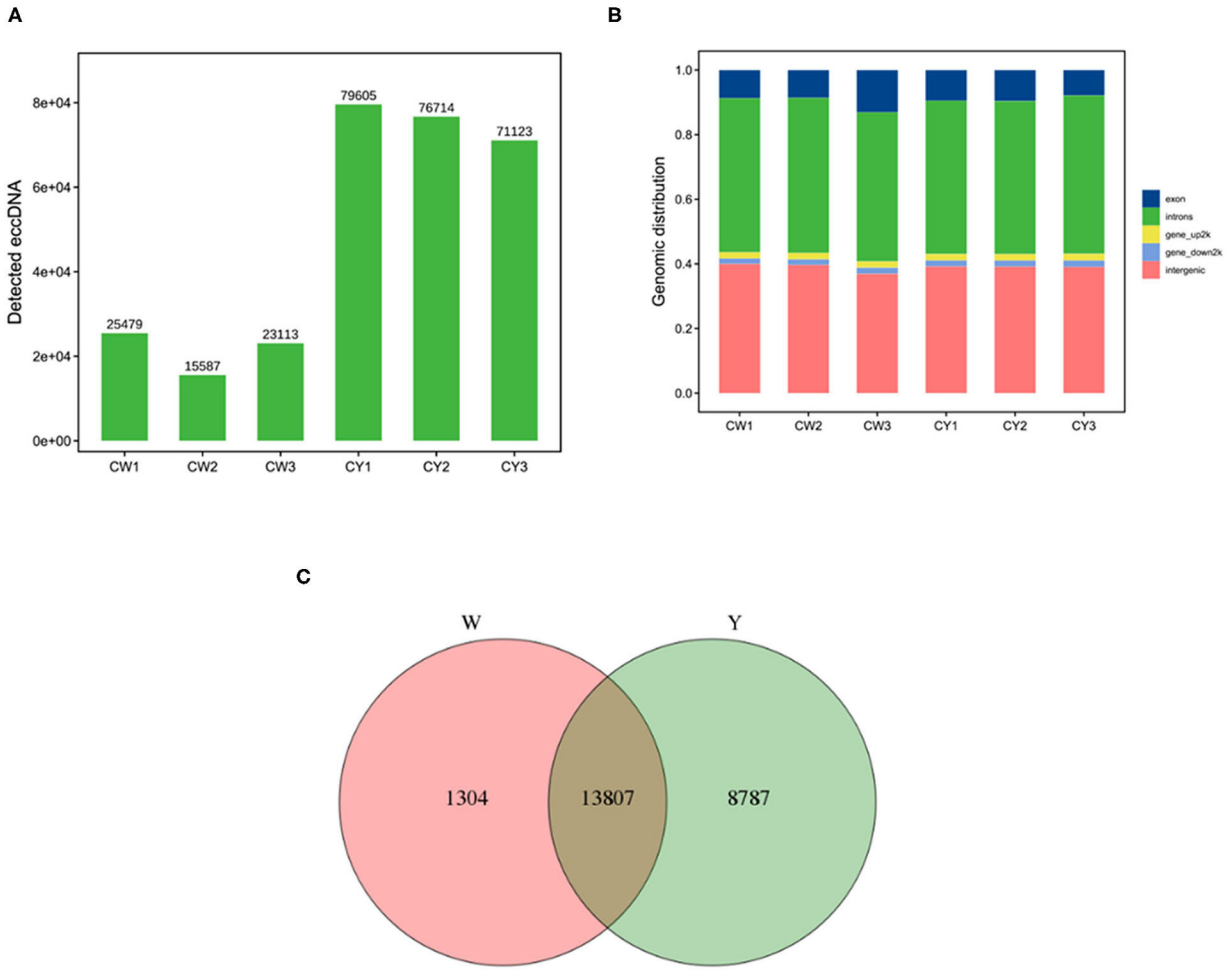


FIGURE 1
Identification of eccDNA and its derived genes. **(A)** Identification of eccDNA in Wei (CW) and Large White (CY) pigs. **(B)** Distribution of eccDNA. **(C)** Venn analysis of eccDNA derived genes in Wei and Large White pigs.

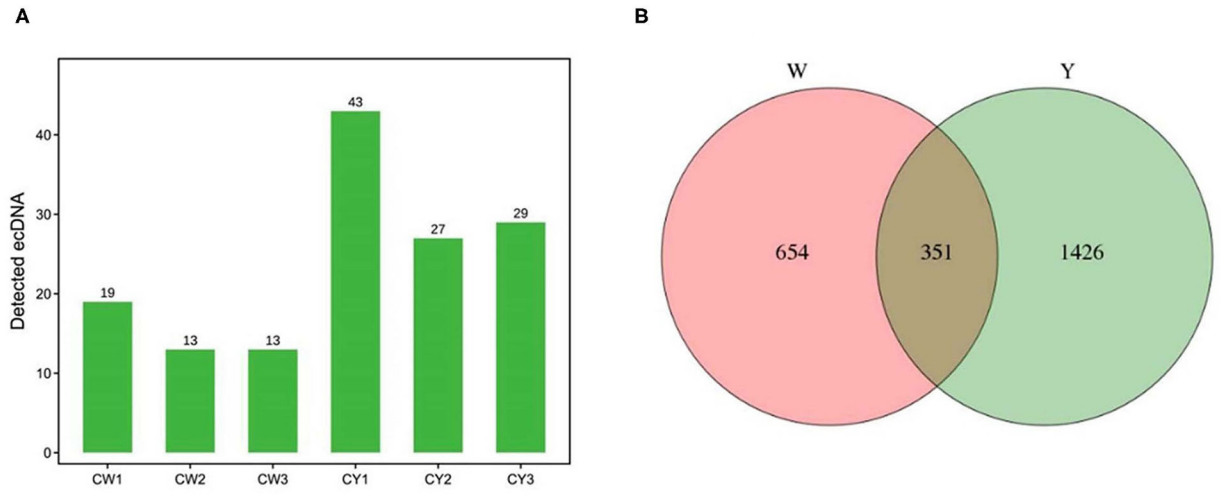


FIGURE 2
Identification of ecDNA and its derived genes. **(A)** Identification of ecDNA in Wei (CW) and Large White (CY) pigs. **(B)** Venn analysis of ecDNA derived genes in Wei and Large White pigs.

Results

Summary of high-throughput sequencing

Sequencing yielded 39,686,953,656–58,411,217,258 and 53,824,168,657–58,311,810,737 clean data for WP and LP, respectively (Table 1). The average Q30 values for WP and LP were 93.69 and 92.4%, respectively. An average of 80.46% (WP) and 81.44% (LP) of reads were mapped along the *Sus scrofa* reference genome (Table 1). The average GC content of WP and LP was 49.78 and 54%, respectively.

Identification of eccDNA and its derived genes

As shown in Figure 1A, sequencing yielded 15,587–25,479 and 71,123–79,605 eccDNAs in the ears of WP and LP, respectively. Most eccDNAs belonged to introns and intergenic regions (Figure 1B). There were 15,111 and 22,594 eccDNA-derived genes in the WP and LP, respectively, and 13,807 eccDNA-derived genes were common in the ears of both pig types (Figure 1C).

Identification of ecDNA and its derived genes

Figure 2A shows that sequencing yielded 13–19 and 27–43 ecDNAs in the ears of WP and LP, respectively. There were 1,005 and 1,777 ecDNA-derived genes in WP and LP, respectively, and 351 ecDNA-derived genes were common in the ears of both pig types (Figure 2B).

GO and KEGG analyses of eccDNA-derived genes in WP and LP

As shown in Table 2, the GO enriched terms of eccDNA-derived genes including biological processes (cellular component organization, cellular component organization or biogenesis, metabolic process, etc.), cellular components (cell, cell part,

TABLE 2 GO analysis of common eccDNA derived genes in Wei and Large White pigs.

GO ID	*q*-value	Class	Description
GO:0016043	1.81E-62	Biological process	Cellular component organization
GO:0071840	8.28E-57		Cellular component organization or biogenesis
GO:0008152	1.03E-52		Metabolic process
GO:0071704	3.80E-45		Organic substance metabolic process
GO:0048856	4.21E-43		Anatomical structure development
GO:0044237	1.98E-42		Cellular metabolic process
GO:0032502	5.80E-42		Developmental process
GO:0051234	1.45E-37		Establishment of localization
GO:0007275	1.78E-37		Multicellular organism development
GO:0048518	2.32E-39		Positive regulation of biological process
GO:0005623	7.20E-272	Cellular component	Cell
GO:0044464	7.20E-272		Cell part
GO:0005622	5.30E-208		Intracellular
GO:0044424	5.30E-208		Intracellular part
GO:0043226	1.02E-146		Organelle
GO:0005488	1.08E-141		Binding
GO:0005737	1.90E-133		Cytoplasm
GO:0043229	4.12E-132		Intracellular organelle
GO:0043227	6.93E-104		Membrane-bounded organelle
GO:0044444	3.68E-92		Cytoplasmic part
GO:0005515	1.87E-106	Molecular function	Protein binding
GO:0003824	2.69E-60		Catalytic activity
GO:0043167	5.55E-58		Ion binding
GO:0043168	4.53E-48		Anion binding
GO:0032559	7.20E-44		Adenyl ribonucleotide binding
GO:0030554	1.69E-43		Adenyl nucleotide binding
GO:0005524	6.89E-42		ATP binding
GO:0008144	2.02E-40		Drug binding
GO:0000166	4.15E-40		Nucleotide binding
GO:1901265	4.15E-40		Nucleoside phosphate binding

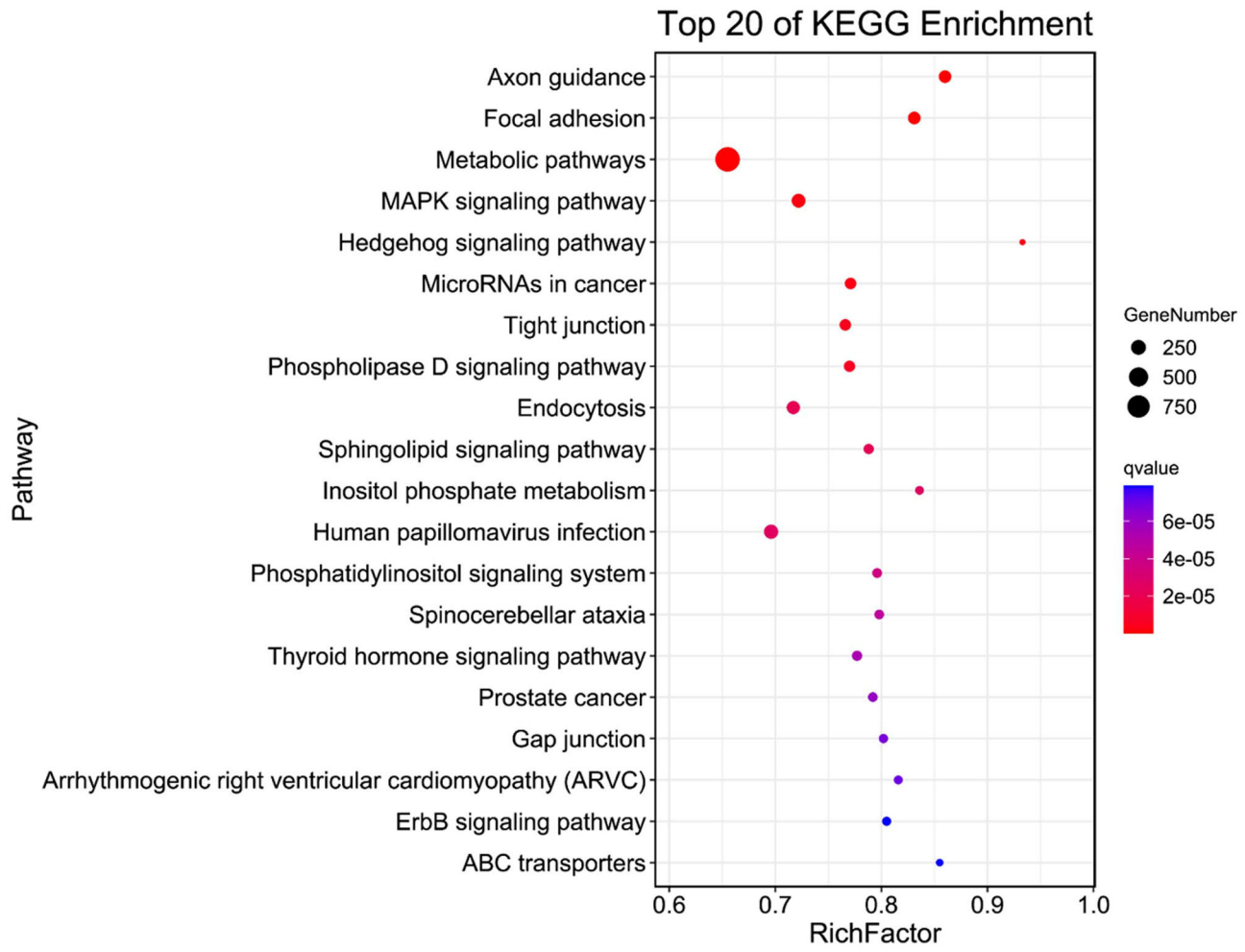


FIGURE 3
KEGG analysis of common eccDNA derived genes in Wei and Large White pigs.

intracellular etc.), and molecular functions (protein binding, catalytic activity, ion binding etc.) in the WP and LP.

Figure 3 shows the top 20 most significant KEGG pathways, including axon guidance, focal adhesion, metabolic pathways, MAPK signaling pathway, Hedgehog signaling pathway, microRNAs in cancer, tight junction, phospholipase D signaling pathway, endocytosis, and sphingolipid signaling pathway in the WP and LP.

GO and KEGG analyses of ecDNA-derived genes in WP and LP

Table 3 shows the GO enriched terms of ecDNA-derived genes, including biological processes (G protein-coupled receptor signaling pathway), cellular components (membrane, integral component of membrane, intrinsic component of membrane, etc.), and molecular functions (olfactory receptor activity, N-acyltransferase activity, G protein-coupled receptor activity, etc.) in the WP and LP.

As shown in Table 4, the top 3 most significant KEGG pathways including olfactory transduction, B cell receptor signaling pathway, and chemical carcinogenesis in the WP and LP.

Differential abundance of eccDNA and ecDNA

As shown in Table 5, eccDNA00044301 was lower abundance in WP compared with that in LP. Furthermore, the gene derived from eccDNA00044301 is a rhomboid 5 homolog 2 (RHBDF2). The ecDNA00000060 was higher abundance in WP compared with that in LP.

Discussion

Generally, most DNA is linear in eukaryote. Nevertheless, some unconventional DNAs, existed in extrachromosomal region, are found to be circular (2, 17). The extrachromosomal cDNA is highly correlated with disease occurrence (7, 18). Pigs and humans are not only highly similar in disease occurrence and physiological characteristics, but also have a high homology in genome and chromosome structure; therefore, pigs have been used as an important mammalian model in human research (19). Many studies have used high-throughput sequencing to study the differences in mRNA, microRNA, and lncRNA in different breeds of pigs, but to

TABLE 3 GO analysis of common ecDNA derived genes in Wei and Large White pigs.

GO ID	*q*-value	Class	Description
GO:0007186	0.004718	Biological process	G protein-coupled receptor signaling pathway
GO:0016020	0.034356	Cellular component	Membrane
GO:0016021	0.034356		Integral component of membrane
GO:0031224	0.035249		Intrinsic component of membrane
GO:0004984	3.22E-05	Molecular function	Olfactory receptor activity
GO:0016410	0.00047		N-acyltransferase activity
GO:0004930	0.00047		G protein-coupled receptor activity
GO:0004888	0.002816		Transmembrane signaling receptor activity
GO:0038023	0.003332		Signaling receptor activity
GO:0060089	0.003332		Molecular transducer activity
GO:0008080	0.003332		N-acetyltransferase activity
GO:0015019	0.006628		Heparan-alpha-glucosaminide N-acetyltransferase activity
GO:0033754	0.006628		Indoleamine 2,3-dioxygenase activity
GO:0016407	0.009387		Acetyltransferase activity
GO:0016747	0.009387		Transferase activity, transferring acyl groups other than amino-acyl groups
GO:0004060	0.013001		Arylamine N-acetyltransferase activity
GO:0008392	0.013001		Arachidonic acid epoxygenase activity
GO:0005452	0.013001		Inorganic anion exchanger activity
GO:0016746	0.018849		Transferase activity, transferring acyl groups
GO:0008391	0.022023		Arachidonic acid monooxygenase activity

TABLE 4 KEGG analysis of common ecDNA derived genes in Wei and Large White pigs.

Pathway	Class	KEGG ID	*q*-value
Olfactory transduction	Organismal systems	ko04740	1.60E-09
B cell receptor signaling pathway	Organismal systems	ko04662	0.0141
Chemical carcinogenesis	Human diseases	ko05204	00179

TABLE 5 Differential abundance of eccDNA and ecDNA in Wei and Large White pigs.

ID	Log_2FC	*P*-value	Type	Annotation	Symbol
EccDNA00044301	−19.091	0.04913	Introns	ENSSSCG00000023362	RHBDF2
EcDNA00000060	5.64542	0.03727	Intergenic	Not available	Not available

the best of our knowledge, there are no previous reports on the high-throughput sequencing of circular DNA in pigs (20–22). In this study, high-throughput sequencing was used to investigate their differences in circular DNA between fatty and lean pig breeds to lay a foundation for their use as human research models and in livestock breeding.

DNA is mainly located in the nucleus of the cells of humans, animals, and plants. However, in special cases, excluding exogenous nucleic acid invasions such as exogenous virus infection, some DNA particles are present outside the chromosome (23, 24). These extrachromosomal DNA particles can be linear or circular. Initially, cDNA was identified in human tumor cells and was thought to be associated with tumor heterogeneity and drug resistance (25, 26). Further studies showed that eccDNA was present in yeast, tumor samples, and cancer cell lines (27, 28). Møller et al. (16) suggested that cDNAs are found in the somatic tissue of healthy humans (16). Similar results were obtained in this study. About > 80% clean reads were mapped to *S. scrofa* genome, and many eccDNAs and ecDNAs were isolated from the ear samples of WP and LP. Most eccDNAs belonged to the intron and intergenic regions. These results demonstrated that the eccDNA and ecDNA were not only existed in the tumor cells, but also existed in the somatic tissue.

EccDNA and ecDNA usually carry partial or complete genes and functional elements, and participate in aging, drug resistance, and tumors (27, 29, 30). In the GO enrichment analysis, the eccDNA-

and ecDNA-derived genes were involved in some pathways, including metabolic process, cell, cell part, olfactory receptor activity, and catalytic activity. The KEGG pathway, detected in terms of the number of genes and q value, was analyzed to further discover the potential biological functions of eccDNA and ecDNA-derived genes in the present study. The most significant KEGG pathways of eccDNA-derived genes were annotated as lipid metabolism, folding, sorting and degradation, signal transduction, transport and catabolism, endocrine system, and infectious diseases. Furthermore, the most significant KEGG pathways of ecDNA-derived genes were annotated as organismal systems and human diseases. These results suggest that eccDNA and ecDNA may be important in regulating metabolism and disease occurrence (18, 31). In particular, KEGG pathway annotation of human diseases (e.g., cancers, infectious diseases, endocrine and metabolic diseases, and immune diseases) suggested that WP and LP have significant potential to construct biomedical disease models.

To further investigate the differences in cDNA, the differential abundance of eccDNA and ecDNA was analyzed in WP and LP. EccDNA00044301 was higher abundance in LP than that in WP. The gene derived from EccDNA00044301 is RHBDF2. This gene is important for iRhoms related to physiological targets, development, disease, and targeting therapeutic opportunities (32). The iRhoms have crucial functions in neurological disorders (Alzheimer's and Parkinson's), inflammation, wound healing, skin diseases, and cancer (33). In addition, RHBDF2 participates in cellular oxidative stress and inflammatory reactions (34).

Conclusion

EccDNAs and ecDNAs are common in WP and LP. These cDNAs are large enough to carry one or several partial or complete genes. Both WP and LP can potentially be used to construct human biomedical disease models. Besides, there were existed differential abundance of eccDNA (EccDNA00044301) and ecDNA (EcDNA00000060) between both breeds. Future studies should focus on the functional verification of EccDNA00044301 and EcDNA00000060.

Data availability statement

The data presented in the study are deposited in the China National Center for Bioinformation (CNCB), Genome Sequence Archive (GAS) repository, accession number CRA009400.

Ethics statement

The animal study was reviewed and approved by Animal Care and Use Committee of Anhui Science and Technology University.

Author contributions

HH and AW design the study and directed study implementation. WZ managed the fieldwork. HH and WZ performed laboratory work. HH managed data. AW and KW analyzed data. HH, AW, and WZ developed the first draft of the paper. All authors reviewed and edited drafts of the manuscript and approved the final version.

Funding

This research was supported by the Special Fund for Anhui Agriculture Research System (AGCYJSTX-05-14/15), Key Research and Development Projects of Anhui Province (202004f06020048), the University Synergy Innovation Program of Anhui Province (GXXT-2021-055), and Scientific Research Project of Anhui Science and Technology University (2021zrzd04).

Conflict of interest

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

Publisher's note

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

References

1. Liao Z, Jiang W, Ye L, Li T, Yu X, Liu L. Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *BBA Rev Cancer.* (2020) 1874:188392. doi: 10.1016/j.bbcan.2020.188392

2. Sin STK, Jiang P, Deng J, Lu J, Cheng SH, Dutta A, et al. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc Natl Acad Sci USA.* (2020) 117:1658–65. doi: 10.1073/pnas.1914949117

3. Hung KL, Mischel PS, Chang HY. Gene regulation on extrachromosomal DNA. *Nat Struct Mol Biol.* (2022) 29:736–44. doi: 10.1038/s41594-022-00806-7

4. Hung KL, Yost KE, Xie L, Shi Q, Helmsauer K, Luebeck J, et al. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature.* (2021) 600:731–6. doi: 10.1038/s41586-021-04116-8

5. Ling X, Han Y, Meng J, Zhong B, Chen J, Zhang H, et al. Small extrachromosomal circular DNA (eccDNA): Major functions in evolution and cancer. *Mol Cancer.* (2021) 20:1–15. doi: 10.1186/s12943-021-01413-8

6. Li R, Wang Y, Li J, Zhou X. Extrachromosomal circular DNA (eccDNA): An emerging star in cancer. *Biomark Res.* (2022) 10:1–13. doi: 10.1186/s40364-022-00399-9

7. Yang L, Jia R, Ge T, Ge S, Zhuang A, Chai P, et al. Extrachromosomal circular DNA: Biogenesis, structure, functions and diseases. *Sig Transduct Target Ther.* (2022) 7:1–21. doi: 10.1038/s41392-022-01176-8

8. Monger XC, Gilbert AA, Saucier L, Vincent AT. Antibiotic resistance: From pig to meat. *Antibiotics.* (2021) 10:1209. doi: 10.3390/antibiotics10101209

9. Natonek-Wiśniewska M, Piestrzynska-Kajtoch A, Koseniuk A, Koseniuk A, Krzyścin P. Current analytical methods and research trends are used to identify

domestic pig and wild boar DNA in Meat and Meat Products. *Genes.* (2022) 13:1825. doi: 10.3390/genes13101825

10. Bassols A, Costa, C, Eckersall PD, Osada J, Sabrià J, Tibau J. The pig as an animal model for human pathologies: A proteomics perspective. *Prot Clin Appl.* (2014) 8:715–31. doi: 10.1002/prca.201300099

11. Heinritz S, Mosenthin R, Weiss E. Use of pigs as a potential model for research into dietary modulation of the human gut microbiota. *Nutr Res Rev.* (2013) 26:191–209. doi: 10.1017/S0954422413000152

12. Meurens F, Summerfield A, Nauwynck H, Saif L, Gerdts V. The pig: a model for human infectious diseases. *Trends Microbiol.* (2012) 20:50–7. doi: 10.1016/j.tim.2011.11.002

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

14. Xu J, Wang C, Jin E, Gu Y, Li S, Li Q. Identification of differentially expressed genes in longissimus dorsi muscle between Wei and Yorkshire pigs using RNA sequencing. *Genes Genom.* (2018) 40:413–21. doi: 10.1007/s13258-017-0643-3

15. Zhang D, Wu W, Huang X, Xu K, Zheng C, Zhang J. Comparative analysis of gene expression profiles in differentiated subcutaneous adipocytes between Jiaxing Black and Large White pigs. *BMC Genomics.* (2021) 22:61. doi: 10.1186/s12864-020-07361-9

16. Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun.* (2018) 9:1–12. doi: 10.1038/s41467-018-03369-8

17. Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature.* (2019) 575:699–703. doi: 10.1038/s41586-019-1763-5 s

18. Qiu GH, Zheng X, Fu M, Huang C, Yang X. The decreased exclusion of nuclear eccDNA: From molecular and subcellular levels to human aging and age-related diseases. *Ageing Res Rev.* (2021) 67:101306. doi: 10.1016/j.arr.2021.101306

19. Schoenly KG, Tarone AM, Villet MH. Pigs vs. people: The use of pigs as analogues for humans in forensic entomology and taphonomy research. *Int J Legal Med.* (2020) 134:793–810. doi: 10.1007/s00414-019-02074-5

20. Liu X, Liu K, Shan B, Wei S, Li D, Han H, et al. A genome-wide landscape of mRNAs, lncRNAs, and circRNAs during subcutaneous adipogenesis in pigs. *J Animal Sci Biotechnol.* (2018) 9:1–13. doi: 10.1186/s40104-018-0292-7

21. Zhang X, Cai S, Chen L, Yuan R, Nie Y, Ding S, et al. Integrated miRNA–mRNA transcriptomic analysis reveals epigenetic-mediated embryonic muscle growth differences between Wuzhishan and Landrace pigs. *J Anim Sci.* (2019) 97:1967–78. doi: 10.1093/jas/skz091

22. Miao Z, Wang S, Wang Y, Wei P, Khan MA, Zhang J, et al. Comparison of microRNAs in the intramuscular adipose tissue from Jinhua and Landrace pigs. *J Cell Biochem*. (2019) 120:192–200. doi: 10.1002/jcb.27298

23. Gu X, Yu A, Chai P, Ge S, Fan X. Novel insights into extrachromosomal DNA: Redefining the onco-drivers of tumor progression. *J Exp Clin Cancer Res.* (2020) 39:1–10. doi: 10.1186/s13046-020-01726-4

24. Wu P, Liu Y, Zhou R, Liu L, Zeng H, Xiong F, et al. Extrachromosomal circular DNA: A new target in cancer. *Front Oncol.* (2022) 12:814504. doi: 10.3389/fonc.2022.814504

25. Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature.* (2017) 543:122–5. doi: 10.1038/nature21356

26. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer.* (2019) 19:283–8. doi: 10.1038/s41568-019-0128-6

27. Zhou T, Ma S, Zhao Y, Guo D, Wang H, Kuang M, et al. Identification and characterization of extrachromosomal circular DNA in alcohol induced osteonecrosis of femoral head. *Front Genet.* (2022) 13:918379. doi: 10.3389/fgene.2022.918379

28. Arrey G, Keating ST, Regenberg B. A unifying model for extrachromosomal circular DNA load in eukaryotic cells. *Semin Cell Dev Biol.* (2022) 128:40–50. doi: 10.1016/j.semcdb.2022.03.002

29. Møller H, Parsons L, Jørgensen TS, Botstein D, Regenberg B. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci USA.* (2015) 112:3114–22. doi: 10.1073/pnas.1508825112

30. Lin C, Chen Y, Zhang F, Liu B, Xie C, Song Y. Encoding gene RAB3B exists in linear chromosomal and circular extrachromosomal DNA and contributes to cisplatin resistance of hypopharyngeal squamous cell carcinoma *via* inducing autophagy. *Cell Death Dis.* (2022) 13:171. doi: 10.1038/s41419-022-04627-w

31. Zhao Y, Yu L, Zhang S, Su X, Zhou X. Extrachromosomal circular DNA: Current status and future prospects. *Elife.* (2022) 11:e81412. doi: 10.7554/eLife.81412

32. Burzenski LM, Low BE, Kohar V, Shultz LD, Wiles MV, Hosur V. Inactive Rhomboid proteins RHBDF1 and RHBDF2 (iRhoms): A decade of research in murine models. *Mamm Genome.* (2021) 32:415–26. doi: 10.1007/s00335-021-09910-3

33. Lichtenthaler SF, O'Hara BF, Blobel CP. iRhoms in the brain–a new frontier? *Cell Cycle.* (2015) 14:3003–4. doi: 10.1080/15384101.2015.1084187

34. Xu MX, Dai XL, Kuang Q, Zhu L, Hu LF, Lou DS. Dysfunctional Rhbdf2 of proopiomelanocortin mitigates ambient particulate matter exposure-induced neurological injury and neuron loss by antagonizing oxidative stress and inflammatory reaction. *J Hazard Mater.* (2020) 400:123158. doi: 10.1016/j.jhazmat.2020.123158

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 05 May 2023
DOI 10.3389/fvets.2023.1125262

OPEN ACCESS

EDITED BY
Sihua Jin,
Anhui Agricultural University, China

REVIEWED BY
Hao Bai,
Yangzhou University, China
Zengkui Lu,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Suli Fang
✉ fangsl@gsau.edu.cn

RECEIVED 16 December 2022
ACCEPTED 13 April 2023
PUBLISHED 05 May 2023

CITATION
Caiye Z, Song S, Li M, Huang X, Luo Y and Fang S (2023) Genome-wide DNA methylation analysis reveals different methylation patterns in Chinese indigenous sheep with different type of tail.
Front. Vet. Sci. 10:1125262.
doi: 10.3389/fvets.2023.1125262

Genome-wide DNA methylation analysis reveals different methylation patterns in Chinese indigenous sheep with different type of tail

Zhu Caiye[1], Shuzhen Song[2], Minna Li[1], Xaioyu Huang[1], Yan Luo[3] and Suli Fang[1]*

[1]College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, Gansu, China, [2]Institute of Animal and Pasture Science and Green Agriculture, Gansu Academy of Agricultural Sciences, Lanzhou, Gansu, China, [3]Gansu Institute of Animal Science and Veterinary Medicine, Pingliang, China

Background: The study was aimed to analyze the difference of genome-wide DNA differential methylation in Lanzhou Large-tailed sheep, Altay sheep and Tibetan sheep, which the typical breeds with different type tails, as to screen the differentially methylated genes (DMGs) that affect the type of tails.

Methods: In this study, three Lanzhou Large-tailed sheep, three Altay sheep and three Tibetan sheep were detected by whole genome bisulfite sequencing (WGBS). The degree of genome-wide DNA methylation, differentially methylated regions (DMRs) and DMGs were analyzed. The candidate genes affecting the tail type of sheep were identified by GO and KEGG pathway enrichment analysis of DMGs.

Results: we identified 68,603 different methylated regions (DMCs) and 75 differentially methylated genes (DMGs) associated with these DMCs. Functional analysis showed that these DMGs were mainly enriched in biological process, cellular component and molecular function, Some of the genes in these pathways are involved in fat metabolism: *NFATC4*, *LPIN2*, *MGAT2* and *MAT2B*.

Conclusion: Our results may help to further understand the epigenetic regulation mechanisms of deposition of fat in the tail of sheep and provide new basic data for the study of local sheep.

KEYWORDS

sheep, tail type, tail fat deposition, DNA methylation, genome

Introduction

Fat deposited on the tail of a sheep is an adversity biological trait, Its role is manifested in two aspects: (1) fat as a form of energy storage, can provide energy for sheep to migrate or survive the cold winter lack of forage, so as to maintain their survival needs (1); (2) in drought and In the famine years, fat at the tail of sheep can be used as a high-energy food for human use (2). Fat-tailed sheep are derived from wild thin-tailed sheep through artificial domestication and natural selection (3). In the process of domestication and selection, the climatic environment and human needs and different selection methods will have an important impact on the

formation of different tail types. Most fat in Lanzhou Large-tailed sheep and Altay sheep is deposited in the tail, which will reduce the deposition of fat in other parts of the carcass, which will affect meat quality (4), moreover, the fat needs more feed than meat production. Therefore, tail fat deposition reduces the benefits of sheep farming, and studying the molecular mechanism of sheep tail fat deposition has important scientific significance and broad application prospects in breeding small-tail low-fat sheep breeds.

In mammals, DNA methylation refers to the supply of methyl donors with S-adenosylmethionine (SAM) under the action of DNA-methyl transferase (DNMT), Transfer its methyl group to the 5th carbon atom of the deoxycytosine ring to form a covalent modification of methylated deoxycytosine (5mC) (5). 5mC generally appears on the cytosine of CpG, CpG sites can account for 5–10% of the mammalian genome. The methylation status of CpG is closely related to gene expression, DNA methylation can inhibit the activity of the accessory gene, and demethylation can induce gene re-expression. Phenotypic differences are not entirely explained by genetic differences, studies have shown that DNA methylation can explain phenotypic differences, such as the phenotypic difference of twins, cloned animals (6–8). DNA methylation regulates the growth and development of adipose tissue mainly by regulating the expression of transcription factors related to adipocyte differentiation, transcription cofactors and genes related to fat metabolism (9). Zhang have shown that methylation of the gene promoter region may inhibit the expression of genes related to fat metabolism, thereby affecting lipid droplet structure and fat deposition (10).

The tail type of sheep is a complex trait, which was affected or interacted by genetic and environmental factors. DNA methylation plays an important role in growth and gene expression in sheep. Mo correlation analysis re studies have shown that DNA methylation can explain phenotypic differences. Chinese indigenous sheep can be classified into fat-tailed, fat-rumped and thin-tailed sheep, of which the typical breeds are Lanzhou Large-tailed sheep, Altay sheep and Tibetan sheep, respectively. To unravel the phenotypic differences among different type tails, in this study, MeDIP-seq was used to perform genome-wide DNA methylation association analysis on three breeds of Lanzhou large-tail sheep, Altay sheep and Tibetan sheep, in order to finding candidate differential methylation regions or sites that potentially affect tail fat deposition, and provide a mechanism for revealing fat deposit mechanisms.

Materials and methods

Ethics statement

All of the animal procedures were performed in strict accordance with the guidelines proposed by the China Council on Animal Care and the Ministry of Agriculture of the People's Republic of China. All of the animal experiments were approved by the Gansu Agricultural University (Lanzhou, China), approval No. GSAU-AEW-2017-0003.

Sample collections

Nine individuals from 3 breeds, all are 1 year old rams, including 3 Lanzhou Large-tailed sheep, 3 Altay sheep and 3 Tibetan sheep, were collected from Yongjing in Gansu Province, Fuhai in Xinjiang Province and Tianzhu in Gansu Province, respectively. All of the individuals were selected with the same sex, weight and age for the same tail type, and there is no genetic relationship between individuals. The tail length, width and circumference of Lanzhou big-tailed sheep are 38, 32, and 106 cm, respectively. The tail length, width and circumference of Altay sheep are 22, 36, and 98 cm respectively, The tail length, width and circumference of Tibetan sheep are 15, 4, and 7 cm, respectively (Figure 1).

Take fresh blood from the jugular vein in the EDTA vein vacuum blood collection tube and store it in the refrigerator at −80°C. Genomic DNA was extracted from the blood samples of 9 individuals using Tiangen blood genomic DNA extraction kit (DP318, Purchased from Beijing Tiangen Biochemical Technology Co., Ltd.). Then through the Nano-Drop spectrophotometer and gel electrophoresis test, reserve after being qualified.

Establishment of MeDIP-seq sequencing library

Genomic DNA ultrasound interrupted to 100–500 bp fragment, DNA fragment end repair add A base at 3'end, connect to sequencing adapter, which were treated by bisulfite conversion with EZ DNA Methylation-Gold kit (Zyom Research Corporation, Irvine, CA, USA). Cut and recycle after desalting and select the library fragment size, qualified libraries were used for computer sequencing. HiSeq2000 (Illumina, San Diego, CA, USA) was used to send to Benatech for library establishment and sequencing.

Bioinformatics analysis

After the sequencing is completed, the sequencing adapters and low-quality data are filtered, the filtered data is compared with the reference genome, and the sequencing result is compared with the reference genome, and the sequence at the only position in the comparison is used for subsequent standard information analysis and personalized analysis. The data filtering criterion are: contains adapter sequences, the proportion of N bases in the sequence exceeds 5% of the total length of the sequence, and the proportion of bases with a quality value less than 20 in the sequence exceeds 50% of the total length of the sequence, if a sequence meets the above three conditions any one of them, then remove this sequence. The data comparison parameters are: BSMAP software is used for comparison, and the comparison mode is map to 2 forward strands, i.e., BSW(++) and BSC(−+).

Calculate the methylation level of the C base

The methylation level of each methylated C base is calculated according to the following formula: Methylation rate at C site = 100* Reads that support methylation/(Reads that support methylation+ Support unmethylated reads). The sites with a sequencing depth greater than 3 are judged as reliable methylation sites, and use Bedtools software to map the methylation sites of different gene elements in each sample (11).

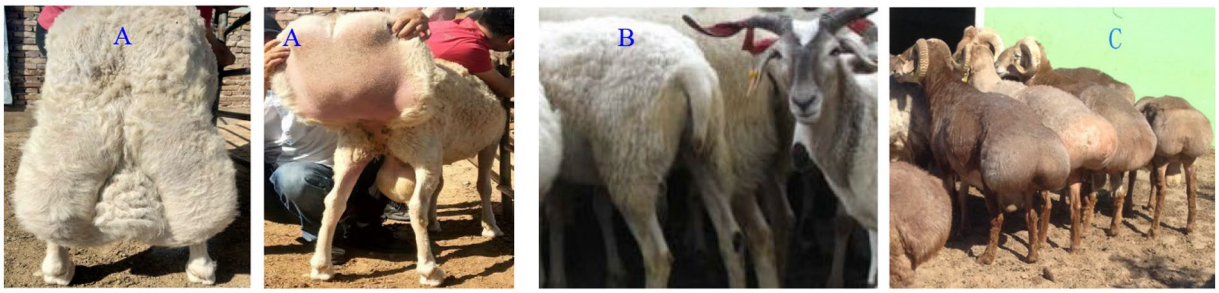


FIGURE 1
(A) Lanzhou Large-tailed sheep with fat-tail. **(B)** Tibetan sheep with thin-tail. **(C)** Altay sheep.

Plotting genome-wide DNA methylation maps

According to the average methylation level of each 100 bp interval, draw the methylation map of the gene body or both sides of the transposable element in the 2Kb region to evaluate the methylation patterns of different genomic regions (12, 13).

Differentially methylated region (DMR) analysis

To determine the methylation types (CpG, CHG and CHH) in the different regions between the 3 species, according to the following strict standards: (1) Find a window containing at least 10 C bases at the same position in the genome of the three samples, (2) Each cytosine site is greater than 10 reads, and each methylated cytosine is greater than 4 reads, (3) The region length ranges from 40 bp–10 Kb, (4) The distance between adjacent methylation sites is less than 200 bp, (5) The average methylation level multiple is greater than 2 and *P*-adjust <0.05. Find the region of differential methylation between different breeds in the adjacent 2 Kb (upstream or downstream), transposon or gene main region.

Differentially methylated gene enrichment analysis

Studies have shown that methylation mainly regulates gene expression through gene promoter regions. Select genes with DMR in the differentially methylated gene promoter (upstream 2 kb region) for gene function GO and signal pathway KEGG enrichment analysis, and select GO/KEGG Term, the results of gene number > 1, enrichment multiple > 2, value of $p<0.05$ (part of the results could not be screened) were plotted.

Results

Date information

The comparison results are shown in Table 1. In Lanzhou Large-tailed sheep, a total of 517,022,996 reads were obtained. The number of reads compared to the sheep genome was 515,987,112, with an average comparison rate of 99.80%. In Altay sheep, total of 523,543,986 reads were obtained and the number of reads compared to the sheep genome was 471,216,836, with an average comparison rate of 90.20%. A total of 529,964, 194 reads were obtained in Tibetan sheep, with a clean rate of 99.80%.

TABLE 1 Results of genome comparison using BWA software.

Sample	Raw reads	Raw bases	Clean reads	Clean rate
Lanzhou Large-tailed sheep	517,022,996	77,553,449,400	515,987,112	99.80%
Altay sheep	523,543,986	78,531,597,900	471,216,836	90.20%
Tibetan sheep	529,964,194	79,494,629,100	528,891,456	99.80%

TABLE 2 Methylation levels in CG, CHG and CHH.

Sample	C	CG	CHG	CHH
Lanzhou Large-tailed sheep	3.62%	64.57%	0.63%	0.65%
Altay sheep	3.67%	65.53%	0.61%	0.63%
Tibetan sheep	3.74%	68.40%	0.59%	0.61%
Average	3.67%	66.16%	0.61%	0.63%

Comparison of genome-wide methylation rates

Distribution of methylation sites in the whole genome

There are three main patterns of DNA methylation at CpG sites (CG, CHG and CHH) in the genome, and the number and composition ratio of these three patterns reflect the characteristics of genome-wide methylation in a given species. The average methylation rate of the whole genome of the three breeds is shown in Table 2. Among the three sheep breeds with different tail types, the CG type has the highest methylation rate, and the average value of the three breeds is 66.16%, which is significantly higher than that of the CHG and CHH types. The distribution of methylation levels in the Table 2 shows that three different tail types of sheep have the highest overall

CG background methylation levels, while CHG and CHH background methylation levels are extremely low or even close to zero.

Analysis of the methylation level of gene functional regional in the whole genome

The analysis showed that CG pattern methylation accounted for most of overall methylation. To be better understanding the distribution trend of DNA methylation in the CG context of genomic regions, we counted the DNA methylation levels on six functional regions and seven transcription elements, including upstream2kb, exon, intron, 3-UTR, downstream 2Kb and intergenic regions. The results showed that the methylation levels of different functional regions and transcription elements were significantly different (Figure 2). Among the functional regions, the 5'UTR region had the lowest methylation level, the 3'UTR and Gene region had the highest average methylation level, and among the transcription elements, the First exon had the lowest methylation level, and the Internal exon, Internal intron, and Last exon regions had the highest methylation level. However, the differences in methylation levels among different samples in the same functional region or transcriptional element were not significant.

Differentially methylated region (DMR) analysis

Find a window containing at least 10C bases at the same position in the genome of the three samples, and use the difference in C methylation levels to find regions with differences in methylation. The statistical results (Table 3) show that the significant DMRs in the CG environment of Altay sheep and Tibetan sheep were 60,631, the significant DMRs in the CG environment between Altay sheep and Lanzhou Large-tailed sheep were 68,603, between Lanzhou Large-tailed sheep and Tibetan sheep were 65,819.

In this study, the DMRs identified above were annotated using the annotation information of functional elements in the Ovis aries genome. Annotate DMR to gene elements (upstream2k, exon, intron, 3-UTR, downstream 2 k, intergenic), and count the distribution of DMR on gene elements. Among them, the most DMRs were found in intron regions (47.2–51.6%), followed by exon regions (4.8–5.3%), and DMRs in other regions were relatively low, only 2.0% of DMRs are distributed in the promoter region, and the methylation rate of the promoter region is lower than the gene body region (Figure 3).

Functional enrichment analysis of differentially methylated genes

DAVID (version 6.8),[1] was used to conduct the GO and KEGG pathway enrichment analyses to investigate the functions of the candidate genes. After this enrichment analysis, followed by the Benjamini correction procedure, there were 20 GO entries were significantly enriched ($p < 0.05$), and 2 KEGG pathways were significantly enriched ($p < 0.05$).

1 https://david.ncifcrf.gov/

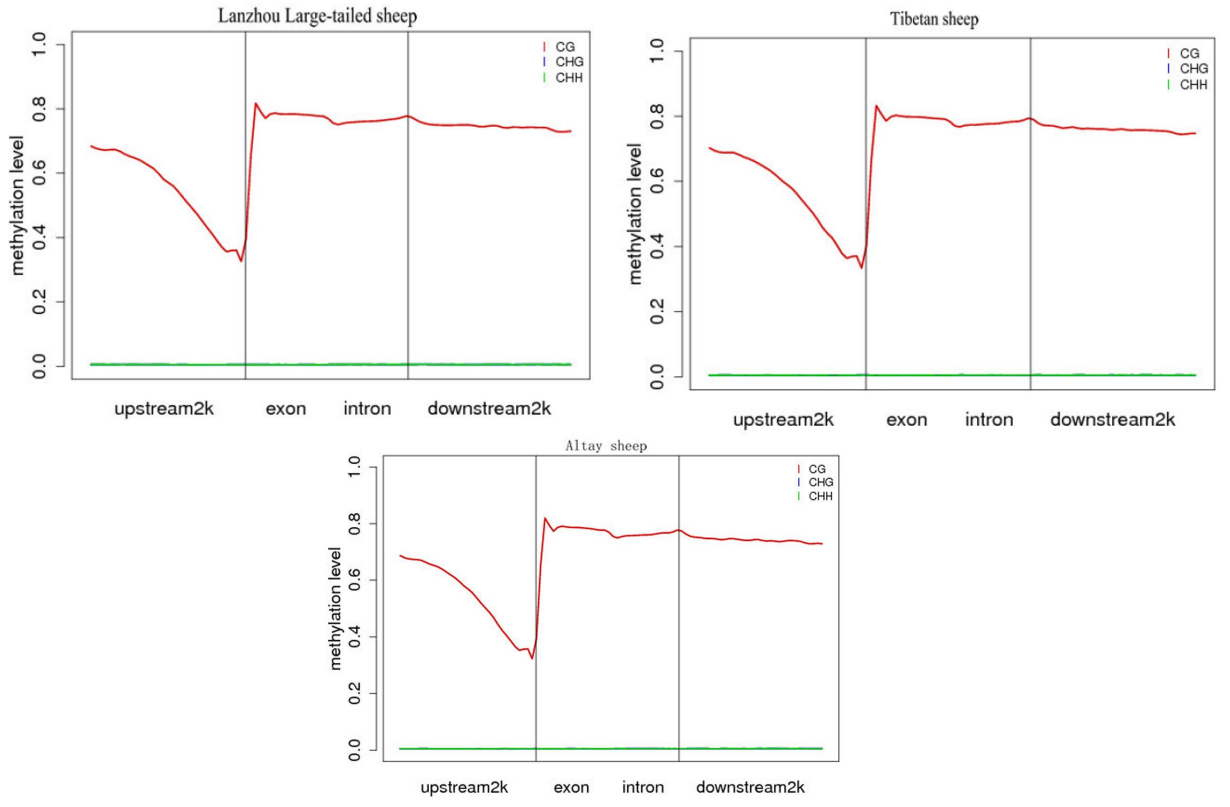


FIGURE 2
Distribution of methylation levels in different functional element regions of the whole country group.

GO results show that these differential genes are mainly related to molecular functions, cell components and biological processes. KEGG enrichment results show that these differential genes are mainly involved in Vitamin B6 metabolism Glycosylphosphatidylinositol (GPI)-anchor biosynthesis, the results are shown in Figure 4.

Combined with genome-wide DMR mapping and DMG function analysis, it was found that most of the genes enriched by GO and KEGG were located in intron regions, and a few were located in Promoter, 5'UTR, and Exon regions. Among the DMR, we found some genes related to fat synthesis and metabolism (Table 4).

Discussion

In the process of sheep domestication, traits that can adapt to the environment and increase energy production will be selected. Under the action of artificial selection, traits related to environmental adaptability and production performance will be further fixed in the genome, and then different breeds will be formed. The epigenetic modification induced by environmental factors will also have a direct impact on some traits of sheep (18). DNA cytosine methylation is an important epigenetic modification, which plays an important role in the regulation of gene expression, development and disease (12).

The main production area of Lanzhou Larger-tailed sheep is located in the suburbs of Lanzhou City. The production area is located in the northwest of the Loess Plateau, with an altitude of 1,500–2,500 m. The winter is longer and the temperature is 9.5°C, the growing season is short, and the sunshine is abundant. The central production area of Altay sheep is in Fuhai County, Xinjiang. The production area is located in the middle mountain belt of the Altai Mountains, with an altitude of 350–3,200 m. It belongs to a typical continental climate. The annual average temperature is 4°C, the extreme minimum temperature is −42.7°C, and the frost-free period is 147 days. The annual snow cover period is 200–250 days. Tibetan sheep are native to the Qinghai-Tibet Plateau and are mainly distributed in Tibet, Qinghai and Gansu. The average altitude is 4,000 meters. It is known as the "roof of the world." The climate is characterized by long sunshine, strong radiation, low temperature, large temperature difference, and winter The spring is dry, the oxygen content is low, and the annual average temperature is −2.8-11.9°C

TABLE 3 Number and length of DMR.

Group	Number of DMR	Number of cytosine	Length of DMR region
ALT108_vs_LZL13	68,603	815,809	27,384,493
ALT108_vs_Z204	60,631	722,624	24,297,767
LZL13_vs_Z204	65,819	776,870	26,097,332

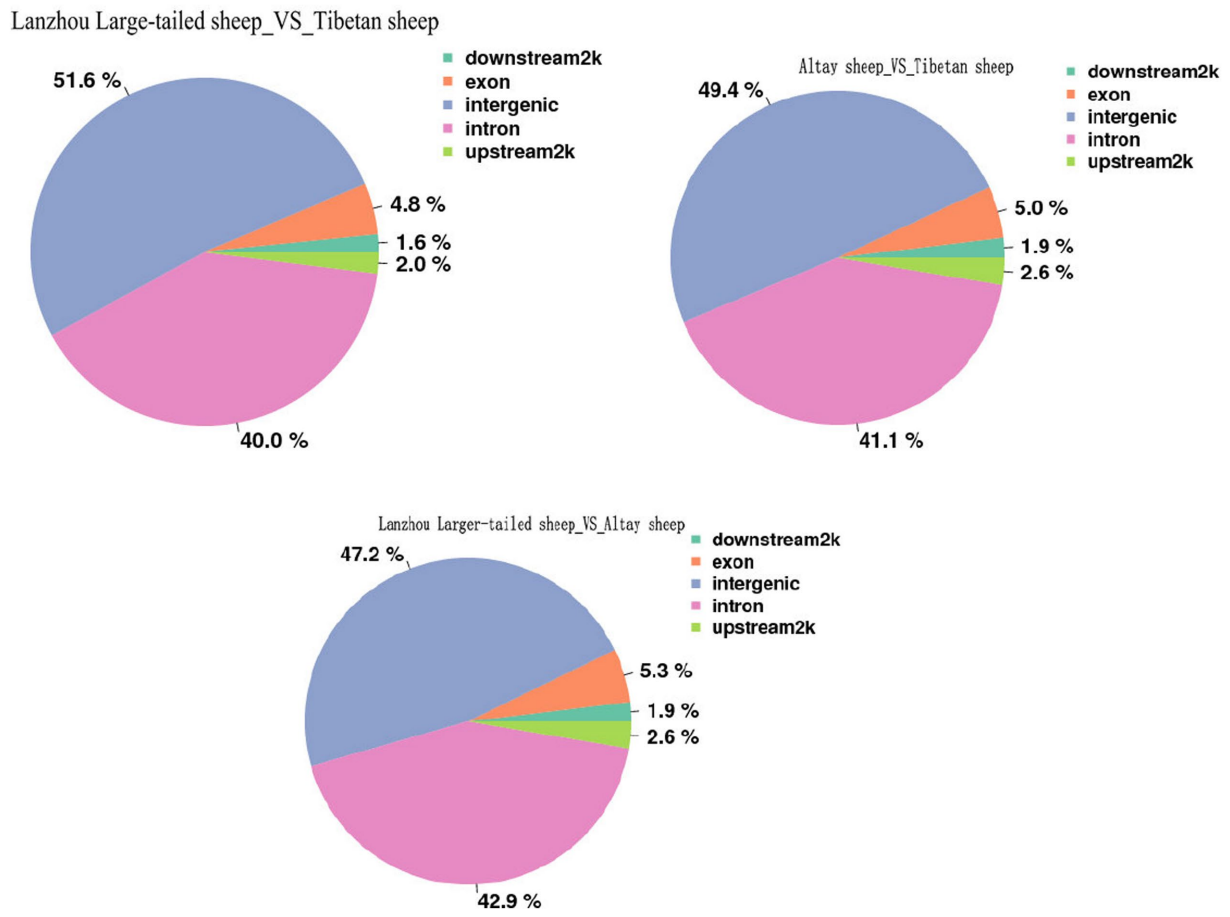


FIGURE 3
Distribution of DMR in different functional element areas.

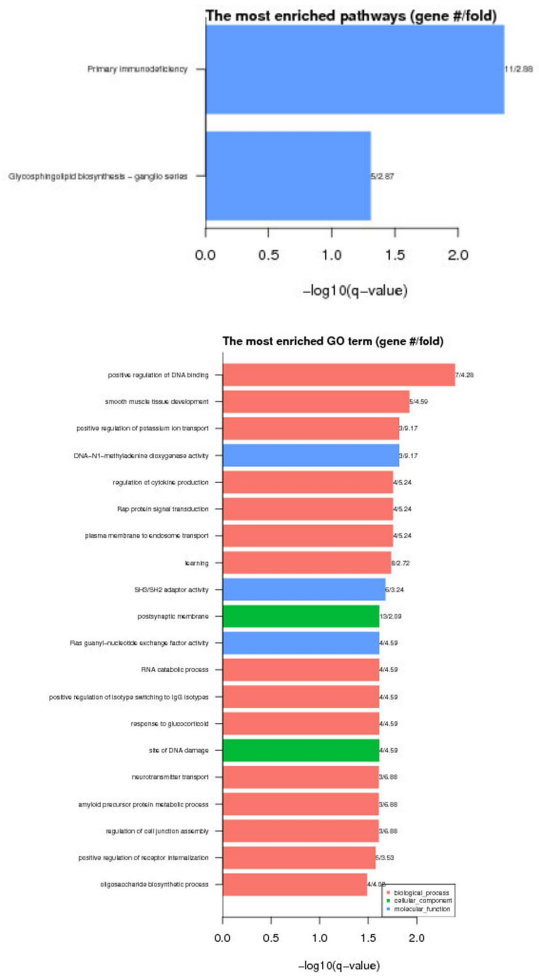


FIGURE 4
Functional enrichment results of differentially methylated genes.

TABLE 4 Fat synthesis and metabolism genes.

Gene name	*p* value	Functional description
NFATC4	0.022	Intramuscular fat deposition (14)
LPIN2	0.024	Related to fatty liver (15)
MGAT2	0.042	Fat deposition (16)
MAT2B	0.033	Regulates fat metabolism (17)

(19). Numerous studies have shown that the interaction between genes and the environment can be explained by DNA methylation. The environment affects gene expression by affecting the level of DNA methylation, thereby changing the phenotype of organisms. The variety we chose is in a different environment and has different tail types. We will explore whether the environment causes methylation, which in turn causes different tail types.

DNA methylation is one of the main forms of epigenetic regulation, which plays an important role in the regulation of gene expression and the maintenance of genome stability (10, 20–22). DNA methylation is caused by the environment. These breeds we selected are in different environments and have different tail types. We will explore whether the environment causes methylation, which in turn causes different tail types. Previous studies have shown that methylation of the CpG site in the promoter region of a specific gene is involved in lipid metabolism (23).

In this study, we performed a genome wide DNA methylation analysis among Lanzhou big-tail sheep, Altay sheep and Tibetan sheep. The geographical environment of these three varieties is completely different. We identified 2,015,487 CpG sites, the methylation level of the CG type accounted for the largest proportion, while that of the CHH and CHG types was the smallest, even close to 0. A large number of studies have also confirmed that the DNA methylation pattern of mammals is dominated by the CG types (24). Our results found on average 66.16% of methylation occurs in the CG environment, this is consistent with the research result of others (25). It is suggested that the DNA methylation pattern has a certain degree of conservation.

We identified the significant DMRs in the CG environment of Altay sheep and Tibetan sheep were 60,631, the significant DMRs in the CG environment between Altay sheep and Lanzhou Large-tailed sheep were 68,603, between Lanzhou Large-tailed sheep and Tibetan sheep were 65,819. Among them, only 2.0% of DMRs are distributed in the promoter region, most of the DMRs are distributed in the intergenic region and intron region, and the methylation rate of the promoter region is lower than the gene body region. This may be related to the large proportion of intergenic regions and intron regions in the entire genome. The percentages of DMCs annotation within promoter and exon region decreased dramatically, while DMCs annotation within intron and intergenic regions increased when comparing DMCs with CpG sites annotation within genic features. And similarly, the percentages of DMCs annotation within CpG island shores and other regions increased, whereas DMCs annotation with CpG islands decreased. Wang et al. also found a similar trend in their study, which is consistent with this study (26, 27). Among the transcription elements, the methylation rate in the internal exon, internal intron, and last exon regions is the highest overall. Methylation patterns similar to the above are also very common in other mammals (28). Although the number of differential methylation regions screened from these three breeds were differed greatly, that three groups of DMR mainly annotated Intron and Exon regions, respectively. Studies have shown that the DNA methylation of Exons and Introns is relatively stable, showing a weak positive and negative correlation with gene expression (29). Introns, which are closely related to gene expression. However, due to the large scale of differential methylation regions in introns, it is difficult to screen functional differential methylation genes.

Based on functional annotation analysis, we found that these DMGs were mainly enriched in positive regulation of DNA binding, smooth muscle tissue development, DNA-N1-methyladenine dioxygenase activity, et al. These results confirmed our hypothesis that Lanzhou Large-tailed sheep, Altay sheep and Tibetan sheep would have epigenetic differences that reflect their specific phenotypic traits among the three breeds. These DMCs and DMGs contribute to the in-depth analysis of molecular genetic mechanism of sheep with different type of tail, and will be given more attention in epigenetic analysis of complex traits for Lanzhou Large-tailed sheep, Altay sheep and Tibetan sheep.

In this study, four tissue-specific candidate genes were screened, including *NFATC4*, *LPIN2*, *MGAT2* and *MAT2B*. *Lipins* play dual function in lipid metabolism by serving as phosphatidate phosphatase

and transcriptional co-regulators of gene expression, Jiao et al. (15) reported that *Lpin2* and *Lpin3* mRNA expression both showed significant correlations with slaughter and tail traits. Taisuke et al. reported that *MGAT2* may have potential for development into a treatment of obesity and its related metabolic diseases (16). Zhao et al. reported that *MAT2A* promotes porcine adipogenesis by mediating H3K27me3 at Wnt10b locus and repressing Wnt/β-catenin signaling (30).

Conclusion

This study performed epigenome-wide DNA methylation of tail fat tissue in Lanzhou Large-tailed sheep, Altay sheep and Tibetan sheep using MeDIP-seq sequencing technology. The results showed that the methylation pattern of sheep with different type were dominated by CG type and the differential methylation region was mainly located in introns and exons. Four annotated genes were associated with fat metabolism (*NFATC4, LPIN2, MGAT2* and *MAT2B*). Our findings provide new insights into a better understanding of the epigenetic regulation of fat deposition in sheep tail and will be useful to understand tail type traits in sheep, and thus contribute to breeding of thin-tail sheep as molecular marker.

Data availability statement

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

Ethics statement

All of the animal procedures were performed in strict accordance with the guidelines proposed by the China Council on Animal Care and the Ministry of Agriculture of the People's Republic of China. All of the animal experiments were approved by the Gansu Agricultural University (Lanzhou, China), approval No. GSAU-AEW-2017-0003.

Author contributions

ZC conceived and designed the experiments, performed the experiments, and wrote the paper. SS, ML, XH, YL, and SF analyzed the data and contributed reagents, materials, and analysis tools. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of China (grant no. 32060748), Fuxi Young Talents Fund of Gansu Agricultural University (GAUfx-04Y012), the Natural Science Foundation of China (grant no. 31960673), and Discipline Team Project of Gansu Agricultural University (No: GAU-XKTD-2022-20).

Conflict of interest

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

Publisher's note

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

References

1. Ermias E, Yami A, Rege J. Fat deposition in tropical sheep as adaptive attribute to periodic feed fluctuation. *J Anim Breed Genet*. (2002) 119:235–46. doi: 10.1046/j.1439-0388.2002.00344.x

2. Kashan NEJ, Azar GHM, Afzalzadeh A, Salehi A. Growth performance and carcass quality of fattening lambs from fat-tailed and tailed sheep breeds. *Small Rumin Res*. (2005) 60:267–71. doi: 10.1016/j.smallrumres.2005.01.001

3. Lv FH, Peng WF, Ji Y, Zhao YX, Li WR, Liu MJ, et al. Mitogenomic meta-analysis identifies two phases of migration in the history of eastern Eurasian sheep. *Molec Biol Evol*. (2015) 32:2515–33. doi: 10.1093/molbev/msv139

4. Negussie E, Rottmann OJ, Pirchner F, Rege J. Patterns of growth and partitioning of fat depots in tropical fat-tailed Menz and Horro sheep breeds. *Meat Sci*. (2003) 64:491–8. doi: 10.1016/S0309-1740(02)00227-9

5. Huang J, Vieira A. DNA methylation, riboswitches, and transcription factor activity: fundamental mechanisms of gene–nutrient interactions involving vitamins. *Mol Biol Rep*. (2006) 33:253–6. doi: 10.1007/s11033-006-9005-y

6. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. (2005) 102:10604–9. doi: 10.1073/pnas.0500398102

7. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GHT, Wong AHC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet*. (2009) 41:240–5. doi: 10.1038/ng.286

8. Rideout III, Eggan K, Jaenisch R. Nuclear cloning and epigenetic reprogramming of the genome. *Science*. (2001) 293:1093–8. doi: 10.1126/science.1063206

9. Huang X, Chen YQ, Xu GL, Peng SH. DNA methylation in adipose tissue and the development of diabetes and obesity. *Yi Chuan*. (2019, 2019) 41:98–110. doi: 10.16288/j.yczz.18-206

10. Zhang S, Shen L, Xia Y, Yang Q, Li X, Tang G, et al. DNA methylation landscape of fat deposits and fatty acid composition in obese and lean pigs. *Sci Rep*. (2016) 6:35063. doi: 10.1038/srep35063

11. Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033

12. Lister R, Pelizzola M, Dowen R, Hawkins D, Hon G, Tonti-Filippini J, et al. Victor: human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. (2009) 462:315–22. doi: 10.1038/nature08514

13. Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, et al. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol*. (2009) 16:564–71. doi: 10.1038/nsmb.1594

14. Kim HB, Kumar A, Wang L, Liu GH, Harris TE. Lipin 1 represses NFATc4 transcriptional activity in adipocytes to inhibit secretion of inflammatory factors. *Mol Cell Biol*. (2010) 30:3126–39. doi: 10.1128/MCB.01671-09

15. Jiao X-L, Jing J-J, Qiao L-Y, Liu J-H, Li L-A, Zhang J, et al. Ontogenetic expression of Lpin2 and Lpin3 genes and their associations with traits in two breeds of Chinese fat-tailed sheep. *Asian Austr J Anim Sci*. (2016) 29:333–42. doi: 10.5713/ajas.15.0467

16. Mochida T, Take K, Maki T, Nakakariya M, Adachi R, Sato K, et al. Inhibition of MGAT2 modulates fat-induced gut peptide release and fat intake in normal mice and

ameliorates obesity and diabetes in Ob/Ob mice fed on a high-fat diet. *FEBS Open Bio*. (2020) 10:316–26. doi: 10.1002/2211-5463.12778

17. Zhao C, Chen X, Wu W, Wang W, Pang W, Yang G. MAT2B promotes adipogenesis by modulating SAMe levels and activating AKT/ERK pathway during porcine intramuscular preadipocyte differentiation. *Exp Cell Res*. (2016) 344:11–21. doi: 10.1016/j.yexcr.2016.02.019

18. Ge CYJ, Weber C, Sun W, Zhang H, Zhou Y, Cai C, et al. The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science*. (2018) 360:645–8. doi: 10.1126/science.aap8328

19. Zhu C, Li N, Cheng H, Ma Y. Genome wide association study for the identification of genes associated with tail fat deposition in Chinese sheep breeds. *Biol Open*. (2021) 10:bio054932. doi: 10.1242/bio.054932

20. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. (2002) 16:6–21. doi: 10.1101/gad.947102

21. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. (2003) 33:245–54. doi: 10.1038/ng1089

22. Wang M, Bissonnette N, Dudemaine PL, Zhao X, Ibeagha-Awemu EM. Whole genome DNA methylation variations in mammary gland tissues from Holstein Cattle producing milk with various fat and protein contents. *Genes*. (2021) 12:1727. doi: 10.3390/genes12111727

23. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. (2012) 13:484–92. doi: 10.1038/nrg3230

24. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. (2013) 16:1–5. doi: 10.1186/s13059-015-0649-6

25. Tian-fu G, Zhi-yan Z, Dong C, Tian-Xiong Y, Shi-Jun X, Lu-sheng H. High resolution and single base genome-wide methylation variance analysis of muscle of large white pigs with different sexes. *Acta Vet Zootech Sin*. (2018) 49:2326–39. doi: 10.11843/j.issn.0366-6964.2018.11.004

26. Wang K, Ping-Xian WU, Wang SJ, Xiang JI, Chen D, Jiang AA, et al. Epigenome-wide DNA methylation analysis reveals differentially methylation patterns in skeletal muscle between Chinese Chenghua and Qingyu pigs. *J Integr Agric*. (2022) 21:1731–9. doi: 10.1016/S2095-3119(21)63814-5

27. Wang X, Kadarmideen HN. An epigenome-wide DNA methylation map of testis in pigs for study of complex traits. *Front Genet*. (2013) 10. doi: 10.3389/fgene.2019.00405

28. Shao X, Zhang C, Sun MA, Lu X, Xie H. Deciphering the heterogeneity in DNA methylation patterns during stem cell differentiation and reprogramming. *BMC Genomics*. (2014) 15:978. doi: 10.1186/1471-2164-15-978

29. Luo R, Bai C, Yang L, Zheng Z, Su G, Gao G, et al. DNA methylation subpatterns at distinct regulatory regions in human early embryos. *Open Biol*. (2018) 8:180131. doi: 10.1098/rsob.180131

30. Zhao CZ, Wu HG, Qimuge P, WJ L. MAT2A promotes porcine adipogenesis by mediating H3K27me3 at Wnt10b locus and repressing Wnt/beta-catenin signaling. *Biochim Biophys Acta Mol Cell Biol Lipids*. (2018) 1863:132–42. doi: 10.1016/j.bbalip.2017.11.001

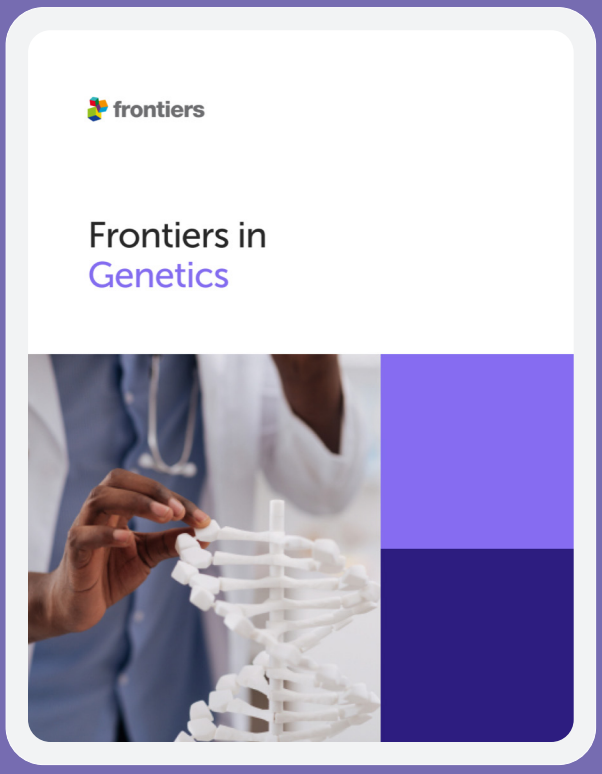
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact