frontiers | Research Topics

Genetic analysis of reproductive traits in livestock

Edited by
Aixin Liang, John S. Davis, Yang Zhou and Hasan Riaz

Published in
Frontiers in Genetics
Frontiers in Veterinary Science

ISSN 1664-8714
ISBN 978-2-83250-784-1
DOI 10.3389/978-2-83250-784-1

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genetic analysis of reproductive traits in livestock

Topic editors

Aixin Liang — Huazhong Agricultural University, China
John S. Davis — University of Nebraska Medical Center, United States
Yang Zhou — Huazhong Agricultural University, China
Hasan Riaz — COMSATS Institute of Information Technology, Pakistan

Citation

Liang, A., Davis, J. S., Zhou, Y., Riaz, H., eds. (2023). *Genetic analysis of reproductive traits in livestock*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-784-1

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1116038

OPEN ACCESS

EDITED AND REVIEWED BY
Johann Sölkner,
University of Natural Resources and Life Sciences Vienna, Austria

*CORRESPONDENCE
John S. Davis,
jsdavis@unmc.edu

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

RECEIVED 05 December 2022
ACCEPTED 15 December 2022
PUBLISHED 04 January 2023

CITATION
Liang A, Zhou Y, Riaz H and Davis JS (2023), Editorial: Genetic analysis of reproductive traits in livestock. *Front. Genet.* 13:1116038. doi: 10.3389/fgene.2022.1116038

Editorial: Genetic analysis of reproductive traits in livestock

Aixin Liang[1], Yang Zhou[1], Hasan Riaz[2] and John S. Davis[3]*

[1]College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]VA Nebraska-Western Iowa Health Care System, Department of biosciences, COMSATS University Islamabad, Sahiwal, Pakistan, [3]Department of Obstetrics and Gynecology, University of Nebraska Medical Center, Omaha, NE, United States

KEYWORDS

reproduction, SNP, GWAS, RNA-seq, molecular pathway

Editorial on the Research Topic
Genetic Analysis of Reproductive traits in livestock

Reproduction is one of the most important traits of livestock to maintain the continuity of the species. Improving the reproductive performance of livestock is of importance and tightly related to the selection intensity and production costs (Schmidt et al., 2019). Candidate gene and high throughput studies have been used to better understand the genetic basis of reproductive traits over the last decades (Óvilo and Valdovinos, 2012). However, the identification and analysis of specific functional and positional variants and molecular regulatory pathways influencing reproductive efficiency remain a challenging task. Therefore, the goal of this Research Topic is to present current knowledge about genetic factors affecting reproduction in animals and present state-of-the-art methods for studying genetic influences on reproductive phenotypes. Based on female or male reproductive traits, the thirteen articles published in this Research Topic cover many different species such as pig, cattle, sheep, rabbit, chicken, duck, and turkey. This article collection shows recent advances, recent technologies, and challenges in livestock reproduction.

High-throughput transcriptome sequencing (RNA-Seq) has become the main approach to identify the key genes related to reproductive traits. The work of Mao et al. describe the use of RNA-seq to screen key genes and lncRNAs that affect the fecundity of pigs. The report highlights an important regulatory role that lncRNA MSTRG.3902.1 may play in rpFSH-induced ovulation by affecting the target gene NR5A2 (nuclear receptor subfamily 5, group A, member 2). Likewise, Du et al. use RNA-seq technology to identify differentially expressed genes in ewe adrenal glands under different photoperiod treatments, and identify several novel mRNA, miRNAs, and lncRNAs, which may regulate sheep seasonal estrus. Using RNA-seq technology, Zhang et al. identified several mRNAs (e.g., *GAMT*, *SOHLH1*, *DMC1*, *MACROD1*, *WNT2B*, *SPIN1*, *CRH*, *TTR*, and *WISP1*) that may have direct or indirect functions in the initiation of puberty, which may provide new insight into the mechanisms that initiate puberty in sheep. Moreover, Ross et al. provided a case study that combined information from multiple expression datasets such as RNA-seq, ISO-seq and CAGE-seq, and identified several genes relevant to fertility in Brahman cattle. They demonstrated tissue-specific expression of the selected genes, allele-specific expression, variation in transcription

start sites, and untranslated regions. The integration of RNA-seq and other sequencing technologies will be a viable alternative to effectively improve the accuracy of candidate gene selection.

In addition to RNA-seq technology, genome-wide association studies (GWASs) and whole-genome sequencing are also widely used to identify key single nucleotide polymorphisms (SNPs) and candidate genes that correspond to reproductive traits. By using GWASs, Mo et al. identified a total of 29 candidate SNPs for seven litter-size traits and four teat-number traits in Bama Xiang pigs. By using GWASs and haplotype-sharing analysis, Xu et al. observed candidate genes and haplotypes that were significantly associated with egg production traits in laying ducks. In addition, Zhang et al. reported on 10 important candidate genes related to bone traits, and two bone-related pathways such as osteoclast differentiation and MAPK (Mitogen Activated Protein Kinase) signaling pathway in laying chicken populations using whole-genome pooled sequencing. By integrated analysis of GWASs and transcriptome data, the study of Shi et al. identified 7 significant SNPs and proposed 28 candidate genes associated with sow milk production, 10 of which were key candidates. Compared to the traditional cumulative model, Makanjuola et al. demonstrated that random regression models using pedigree and genomic information can achieve a higher predictive ability for analyzing longitudinal traits such as fertile eggs set in the incubator (FERT), hatch of fertile (HOF), and hatch of set (HOS) in turkeys.

It has been reported that splicing isoforms may exert distinct functions in reproductive physiological processes, such as progesterone receptor isoforms (Rekawiecki et al., 2011) and prolactin receptor isoforms (Binart et al., 2010). The article of Kern et al. describe four isoforms of preferentially expressed antigen in melanoma Y-linked (PRAMEY) in the bovine testis and spermatozoa. The study implicates that the 58 and 30 kDa PRAMEY isoforms are involved in spermatogenesis, whereas the 13 kDa PRAMEY isoform is responsible for sperm maturation and sperm motility.

Microbial communities in the reproductive tract are involved in the maintenance of host fertility and health (Feng and Liu, 2022). Endometrial inflammation is common in postpartum dairy cows, and alterations in the uterine microbiota are associated with perinatal disease. The study of Kudo et al. revealed that *Trueperella* is present in higher abundance in the uterus and vagina of the endometritis group and is negatively correlated to the abundance of *Lactobacillus*. As mentioned in their article, their findings are helpful for predicting endometritis and developing prevention or treatment strategies.

Myostatin (MSTN) is regarded as a negative regulator of muscle development and regeneration, and natural mutations of *MSTN* result in an obvious double muscle phenotypic effect in cattle, dogs, sheep, and pigs. In this Research Topic, Zheng et al. developed a heritable double muscle buttocks in rabbits *via* *MSTN* mutation using the CRISPR/Cas9 system. This may improve rabbit meat production efficiency and promote the development of the rabbit industry.

In summary, integration of the available technological approaches provides more powerful tools for the identification of novel functional candidate genes, specific genetic variants, and molecular pathways affecting reproductive traits. The CRISPR/Cas9 system is an efficient genome editing tool for the validation of functional genes in relation to reproduction, which may significantly improve reproductive efficiency in livestock.

Author contributions

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

Acknowledgments

We want to thank all the authors for their novel work and the external reviewers for their valuable comments to improve the quality of the articles. We also appreciate the continuous support of the Frontiers staff.

Conflict of interest

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

Publisher's note

References

Binart, N., Bachelot, A., and Bouilly, J. (2010). Impact of prolactin receptor isoforms on reproduction. *Trends Endocrinol. Metab.* 21 (6), 362–368. doi:10.1016/j.tem.2010.01.008

Feng, T., and Liu, Y. (2022). Microorganisms in the reproductive system and probiotic's regulatory effects on reproductive health. *Comput. Struct. Biotechnol. J.* 20, 1541–1555.

Óvilo, C., and Valdovinos, C. (2012). *Animal reproduction in livestock - genetic basis and improvement of reproductive traits*. Oxford ,UK: ©Encyclopedia of Life Support Systems.

Rekawiecki, R., Kowalik, M. K., and Kotwica, J. (2011). Nuclear progesterone receptor isoforms and their functions in the female reproductive tract. *Pol. J. Vet. Sci.* 14 (1), 149–158. doi:10.2478/v10181-011-0024-9

Schmidt, P. I., Campos, G. S., Roso, V. M., Souza, F. R. P., and Boligon, A. A. (2019). Genetic analysis of female reproductive efficiency, scrotal circumference, and growth traits in Nelore cattle. *Theriogenology* 128, 47–53. doi:10.1016/j.theriogenology.2019.01.032

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fgene.2021.724533

Identification of SNPs and Candidate Genes for Milk Production Ability in Yorkshire Pigs

Lijun Shi, Yang Li, Qian Liu, Longchao Zhang, Ligang Wang, Xin Liu, Hongmei Gao, Xinhua Hou, Fuping Zhao, Hua Yan and Lixian Wang**

Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

Sow milk production ability is an important limiting factor impacting suboptimal growth and the survival of piglets. Through pig genetic improvement, litter sizes have been increased. Larger litters need more suckling mammary glands, which results in increased milk from the lactating sow. Hence, there is much significance to exploring sow lactation performance. For milk production ability, it is not practical to directly measure the milk yield, we used litter weight gain (LWG) throughout sow lactation as an indicator. In this study, we estimated the heritability of LWG, namely, 0.18 ± 0.07. We then performed a GWAS, and detected seven significant SNPs, namely, *Sus scrofa* Chromosome (SSC) 2: ASGA0010040 (p = 7.73E-11); SSC2:MARC0029355 (p = 1.30E-08), SSC6: WU_10.2_6_65751151 (p = 1.32E-10), SSC7: MARC0058875 (p = 4.99E-09), SSC10: WU_10.2_10_49571394 (p = 6.79E-08), SSC11: M1GA0014659 (p = 1.19E-07), and SSC15: MARC0042106 (p = 1.16E-07). We performed the distribution of phenotypes corresponding to the genotypes of seven significant SNPs and showed that ASGA0010040, MARC0029355, MARC0058875, WU_10.2_10_49571394, M1GA0014659, and MARC0042106 had extreme phenotypic values that corresponded to the homozygous genotypes, while the intermediate values corresponded to the heterozygous genotypes. We screened for flanking regions ± 200 kb nearby the seven significant SNPs, and identified 38 genes in total. Among them, 28 of the candidates were involved in lactose metabolism, colostrum immunity, milk protein, and milk fat by functional enrichment analysis. Through the combined analysis between 28 candidate genes and transcriptome data of the sow mammary gland, we found nine commons (ANO3, MUC15, DISP3, FBXO6, CLCN6, HLA-DRA, SLA-DRB1, SLA-DQB1, and SLA-DQA1). Furthermore, by comparing the chromosome positions of the candidate genes with the quantitative trait locus (QTLs) as previously reported, a total of 17 genes were found to be within 0.86–94.02 Mb of the reported QTLs for sow milk production ability, in which, NAV2 was found to be located with 0.86 Mb of the QTL region ssc2: 40936355. In conclusion, we identified seven significant SNPs located on SSC2, 6, 7, 10, 11, and 15, and propose 28 candidate genes for the ability to produce milk in Yorkshire pigs, 10 of which were key candidates.

Keywords: Yorkshire pig, litter weight gain, GWAS, SNP, candidate gene

OPEN ACCESS

Edited by:
Hasan Riaz, COMSATS Institute of Information Technology, Pakistan

Reviewed by:
Hugo Oswaldo Toledo-Alvarado, National Autonomous University of Mexico, Mexico
Muhammad Azhar, COMSATS University Islamabad—Sahiwal Campus, Pakistan

****Correspondence:***
Lijun Shi
shilijun01@caas.cn
Lixian Wang
iaswlx@263.net

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

Received: *13 June 2021*
Accepted: *22 September 2021*
Published: *05 October 2021*

Citation:
Shi L, Li Y, Liu Q, Zhang L, Wang L, Liu X, Gao H, Hou X, Zhao F, Yan H and Wang L (2021) Identification of SNPs and Candidate Genes for Milk Production Ability in Yorkshire Pigs. Front. Genet. 12:724533. doi: 10.3389/fgene.2021.724533

INTRODUCTION

The mammary gland is a ubiquitous morphological feature of mammals, and lactation is an essential process in mammalian reproduction, including the secretion of milk from mammary glands. For offspring, depending on milk is a key strategy to the life history of all mammals. During lactation, maintaining body growth and milk production for the dam is necessary, thus energy requirement is high. In the past few decades, genetic and management changes have occurred, and the modern sow is subject to additional challenges. Litter size is one of the most important factors affecting milk production in a sow (Eissen et al., 2000), and piglet survival after birth is negatively affected by increasing litter size (Wang et al., 2017). During this period, the litter size of pigs has increased and will continue as an important goal trait in pig breeding programs around the world (Spötter and Distl, 2006; Baxter et al., 2013). In general, larger litters need more suckle mammary glands, which results in increased milk from the lactating sow (Auldist et al., 1998). The survival of offspring can be enhanced by milk yield, which satisfies the immunological needs of offspring and assists in the endocrine maturation of neonates (Goldman, 2002). In response to greater suckling intensity, sows have to produce more milk to nurse more piglets (Auldist et al., 1998; Revell et al., 1998). Additionally, poor lactation traits lead to early culling, which affects the profitability of commercial producers. Hence, it is of economic importance to improve lactation performance in pigs, and it is necessary to include lactation traits in the breeding goals.

The genetic improvement of sow lactation performance is hindered due to the difficulty of collecting accurate phenotypes. Unlike dairy cattle, it is not possible to directly measure the sow milk yield. Different experimental methods have been proposed to measure pig milk production ability, such as the isotope dilution method (Pettigrew et al., 1987) and the weighsuckle-weigh method (Elsley, 1971). These methods are expensive, complicated, and labor-intensive, and are difficult to be implemented on a routine basis in a commercial herd. A simpler and more straightforward measurement for an increase in body weight of piglets during lactation has been reported and is considered as an indicator trait for milk production ability (Revell et al., 1998; Bergsma et al., 2008). In 2016, DM. Thekkoot et al. estimated the heritability of litter weight gain (LWG) as an indicator of lactation trait in Yorkshire and Landrace sows, namely 0.16–0.22 and 0.12–0.20, respectively (Thekkoot et al., 2016a).

A Genome-Wide Association Study (GWAS) is an effective strategy to examine the underlying genetics of complex traits (Goddard and Hayes, 2009). Many studies have identified candidate markers associated with important economic traits in pigs, such as meat quality (Falker-Gieske et al., 2019) and growth (Zhang et al., 2019). For LWG traits in Yorkshire sow lactation, the GWAS detects two quantitative trait locus (QTLs) on *Sus scrofa* Chromosome (SSC) 7 (126 and 101 Mb) (Thekkoot et al., 2016b).

Until now, there has been little known about the heritability and genomic prediction of sow milk production ability. In this study, we aimed to estimate the heritability of LWG of the sow during lactation, to perform a GWAS for proposing the single nucleotide polymorphisms (SNPs) and candidate genes, and to conduct the combined analysis with the reported swine mammary gland transcriptome data and GWAS data for further insights into the candidates involved in sow milk synthesis.

MATERIALS AND METHODS

Animals and Phenotypic Data

In this study, a total of 985 Yorkshire sows involved in 96 sire families, were recorded between 2019 and 2020 in Shanxi and Liaoning Province, China. These sows were fed with the fodders prescribed by their farms, in which, the regular quarantine inspection was carried out. For each sow, only one production record was performed, and 985 individuals were involved in 1–8 parity.

As it was not practical to directly measure the milk production ability of sows, our study weighed all non-mummified piglets at birth, death, weaning, and at the time of fostering. This allowed us to quantify the exact weight gain of each piglet for each sow. We calculated the LWG for each sow by summing up the increase in weight of all piglets nursed by that sow and considered it as a potential indicator for milk production ability. The formula for calculating LWG was as follow:

$$\begin{aligned}\text{LWG (kg)} = {} & \text{litter weight at weaning} - \text{litter weight at birth} \\ & - \text{litter weight at the time of fostering in} \\ & + \text{litter weight at death} \\ & + \text{litter weight at the time of fostering out}\end{aligned}$$

Genetic Parameters Estimation for LWG

We estimated the genetic parameters of LWG with an animal model. The genetic parameters and estimated breeding values (EBV) were performed by the ASReml package as the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of phenotypic records (LWG of the sow); $\mathbf{b}$ is a vector of fixed effects containing herd by farm and production batch (nine levels), parity (five levels: 1, 2, 3, 4, and 5–8), and days of lactation (three levels: ≤ 18, 19–21, and > 21); $\mathbf{X}$ is a design matrix that associates $\mathbf{b}$ with $\mathbf{y}$; $\mathbf{a}$ is the *vector* of additive genetic effects; Z is the corresponding incidence *matrix*, and $\mathbf{e}$ is the vector of random residual effects. Variances of random effects are defined as $V(a) = \mathbf{G}\sigma_a^2$ for the polygenes and $V(e) = \mathbf{I}\sigma_e^2$ for the residuals, where the $\mathbf{G}$ is the additive genetic relationship matrix, $\mathbf{I}$ is the identity matrix, $V(a)$ is the additive genetic variance, and $V(e)$ is the residual variance. In this study, 985 sows were traced back to four-generation pedigrees to construct the kinship matrix, and a total of 2,415 individuals were included.

Genotyping and Quality Control

Ear samples of the 985 Yorkshire sows were collected in farms. For each ear, DNA was isolated with a commercially available kit, Q1Aamp DNA Mini Kit (QIAGEN, Germany). In total, 985 sows were then genotyped with the GenSeek Genomic Profiler (GGP) Porcine 50K (50,697 SNPs, Illumina, San Diego, CA, United States).

With PLINK (Purcell et al., 2007), we removed the SNPs with minor allele frequencies < 0.01, and a deviation from Hardy-Weinberg equilibrium (HWF) p values < 0.001. A dataset containing 36,871 SNPs and 985 animals was used for further analysis. All SNP positions were annotated based on pig genome assembly *Sscrofa* 11.1. The genotype data used for GWAS was submitted to public repositories, and the DOI was 10.6084/m9.figshare.16545915 (https://figshare.com/s/edda38a1c99aa7ab7ae0).

Genome-Wide Association Study

We utilized the EBV of LWG as the dependent variable to perform GWAS by Fixed and random effect model Circulating Probability Unification (FarmCPU). FarmCPU is a multi-locus model that incorporates multiple markers simultaneously as covariates to partially remove the confounding effect between testing markers and kinship (Liu et al., 2016). A genome-wide Bonferroni correction threshold of 0.05/36,871 (i.e., 1.36E-06) was implemented to correct for multiple testing and assess the significance level for each SNP. The Manhattan and quantile-quantile (Q-Q) plots were drawn by R packages (http://cran.r-project.org/web/packages/gap/index.html).

In addition, we estimated the least square mean of sow LWG phenotypes for homozygous and heterozygous genotypes of the seven significant SNPs with standard error (SE) by SAS9.2 (SAS Institute, Cary, NC, United States).

Gene Contents and Functional Annotation

We used the BioMart in Ensembl database to retrieve candidate genes within 200 kb (Zhao et al., 2011) of significant SNPs based on the pig reference genome (Sscrofa11.1). To provide insight into the functional enrichment of candidate genes identified in this study, we performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis with the KOBAS (http://kobas.cbi.pku.edu.cn/kobas3/genelist/) (Xie et al., 2011).

Combined Analysis With the Reported Transcriptome and GWAS Data

To further confirm the key candidates, we performed the combined analysis between the results of this study and reported transcriptome research of the sow mammary gland (Palombo et al., 2018).

Based on the gene location information in the Ensembl database (http://asia.ensembl.org/index.html) and reported GWAS, it was considered that the candidate genes located within 5 Mb to the peak of QTLs in the previous GWAS were promising candidates associated with the ability to produce milk.

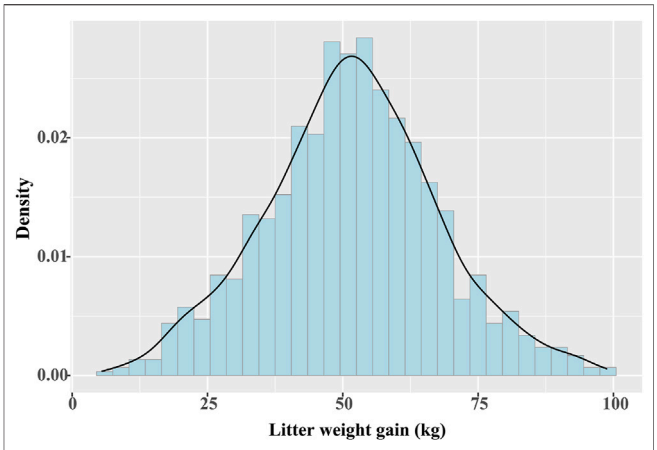


FIGURE 1 | Distribution of milk production ability for 985 Yorkshire pigs.

RESULTS

Descriptive Statistics and Heritability of LWG Trait

For 985 Yorkshire pigs, the average days of lactation were 19.13. We calculated the descriptive statistics of LWG throughout lactation: number sows ~ 985, mean ~ 51.65 kg, standard deviation ~ 16.05, maximum ~ 98.74 kg, and minimum ~ 5.54 kg. **Figure 1** shows the distribution of LWG, which indicated the data was normal.

We estimated the heritability of sow LWG: 0.18 ± 0.07, in which, the estimated additive variance was $V(a) = 25.95 \pm 10.85$, and residual variance was $V(e) = 119.69 \pm 10.30$. Furthermore, we estimated the breeding value and include the results in **Supplementary Table S1**.

GWAS and Identification of Candidate Genes

In this study, a total of 985 sows with the EBVs of LWG and genotypes were used for the GWAS by FarmCPU. The Manhattan and Q-Q plots are shown in **Figures 2A,B**, respectively. Seven genome-wide significant SNPs were identified: ASGA0010040 (p = 7.73E-11) and MARC0029355 (p = 1.30E-08) located on SSC 2, WU_10.2_6_65751151 (p = 1.32E-10) located on SSC6, MARC0058875 (p = 4.99E-09) located on SSC7, WU_10.2_10_49571394 (p = 6.79E-08) located on SSC10, M1GA0014659 (p = 1.19E-07) located on SSC11, and MARC0042106 (p = 1.16E-07) located on SSC15 (**Table 1**).

We performed the distribution of phenotypes for LWG by the genotype of the significant SNPs, the results of which can be seen in **Figure 3**. These data of ASGA0010040, MARC0029355, MARC0058875, WU_10.2_10_49571394, M1GA0014659, and MARC0042106 showed that the extreme phenotypic values corresponded to the homozygous genotypes, while the intermediate values corresponded to the heterozygous genotypes. The least-square mean (± SE) of the LWG by seven

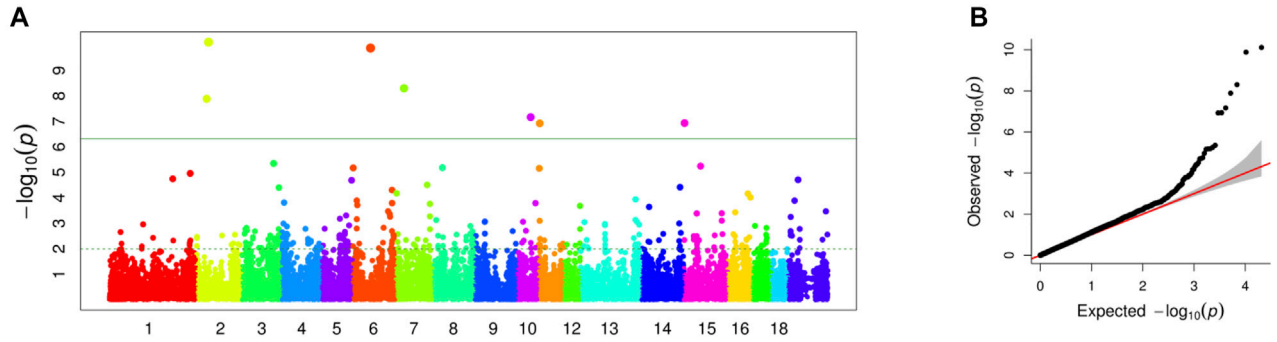


FIGURE 2 | (A) Manhattan plot of association results for sow milk production ability. The green line indicated p = 1.36E-06. **(B)** Quantile-quantile plot of 36,871 SNPs in genome-wide association study for milk production ability.

significant SNPs is shown in **Table 2**, which also presents the genotype and allele frequencies.

Sows that were homozygous AA for ASGA0010040 showed significantly lower LWG than those that were homozygous GG ($p < 0.01$) and heterozygous AG ($p < 0.05$). The homozygous AA for MARC0058875 showed significantly larger LWG than those with homozygous GG ($p < 0.01$) and heterozygous AG ($p < 0.01$). The homozygous AA for M1GA0014659 showed significantly larger LWG than those with homozygous GG ($p < 0.05$). Sows that were homozygous AA for WU_10.2_6_65751151 and AG for MARC0042106 showed significantly larger milk production ability than those that were heterozygous AG ($p < 0.01$) and homozygous AA ($p < 0.05$), respectively. The SNPs MARC0029355 and WU_10.2_10_49571394 were not significant, while the homozygous GG for MARC0029355 and AA for WU_10.2_10_49571394 had obvious larger LWG than those with homozygous AA and GG, respectively. These results further confirmed that the seven SNPs were highly associated with sow milk production ability.

In addition, through screening for flanking regions ± 200 kb nearby seven significant SNPs, a total of 38 genes were identified in SSCs 2, 6, 7, 10, 11, and 15 (**Table 1**).

Functional Analysis of Candidate Genes

To investigate the functions of 38 genes, we performed GO and KEGG pathway analysis by KOBAS. In total, 142 GO and 51 KEGG enrichments were clustered with 28 genes (**Supplementary Table S2**). All these GO and KEGG enrichments were mainly related to cellular components and basic metabolism. In which, many GO and KEGG enrichments were involved in lactose metabolism, colostrum immunity, and milk protein and fat, such as tetrahydrofolate interconversion, thermogenesis, oxytocin signaling pathway, antigen processing and presentation, primary immunodeficiency, immune system process, glycoprotein catabolic process, cGMP-PKG signaling pathway, fat cell differentiation, and MAPK signaling pathway (**Supplementary Table S2**). Additionally, there were also many important metabolism enrichments clustered by these genes, including chloride channel activity, ubiquitin-mediated proteolysis, regulation of cell growth, carbon metabolism, metabolic pathways, ATP binding, and oxidation-reduction process (**Supplementary Table S2**). According to the results of the GO and KEGG enrichments, we considered the 28 genes as candidates for lactose metabolism, colostrum immunity, and milk protein and fat (**Supplementary Table S2**).

Combined Analysis With the Reported Transcriptome of Swine Mammary Gland and GWAS Data of Sow Milk Production Ability

To further detect insights into the association of 28 candidate genes with milk synthesis, we performed the combined analysis between this GWAS and reported transcriptome data (Palombo et al., 2018) to improve the accuracy of the selection of functional genes related to milk production in swine. In total, nine (ANO3, MUC15, DISP3, FBXO6, CLCN6, HLA-DRA, SLA-DRB1, SLA-DQB1, and SLA-DQA1) of 28 candidates were differentially expressed genes at days 14, 10, 6, and 2 before (−) parturition and day 1 after (+) parturition (**Table 3**).

We also compared the chromosome positions of 28 candidates with those of the QTLs from reported GWAS data for milk production ability traits, and a total of 17 genes were found to be within 0.86–94.02 Mb of the reported QTLs for milk yield (**Table 3**). In which, NAV2 was found to be located with 0.86 Mb of QTL region ssc2: 40936355 that was confirmed to have large genetic effects on sow milk yield (**Table 3**).

DISCUSSION

In this study, we estimated the heritability and EBV of LWG and performed a GWAS to screen the candidate genes. We found 28 promising candidates involved in lactose metabolism, colostrum immunity, and milk protein and fat, such as tetrahydrofolate interconversion, primary immunodeficiency, glycoprotein catabolic process, fat cell differentiation, and MAPK signaling pathway.

Our heritability estimates for LWG were 0.18 and were consistent with those reported by DM. Thekkoot, who found the heritability of LWG ranged from 0.16 to 0.22 for Yorkshire and 0.12–0.20 for Landrace sows (Thekkoot et al., 2016a). We

TABLE 1 | Candidate genes associated with milk production ability by genome-wide association study

SNP name	SSC	Position	*p*-value	SNP effect	Candidate genes	Gene symbol	Gene full name	Distance (kb)
ASGA0010040	2	39768974	7.73E-11	0.58	ENSSSCG00000013351	NAV2	Neuron navigator 2	Within
MARC0029355	2	33568298	1.3E-08	−0.71	ENSSSCG00000013338	SLC5A12	Solute carrier family 5 member 12	Within
					ENSSSCG00000013339	ANO3	Anoctamin 3	49.69
					ENSSSCG00000013340	MUC15	Mucin 15, cell surface associated	118.04
WU_10.2_6_65751151	6	71788235	1.32E-10	−0.78	ENSSSCG00000003417	DISP3	Dispatched RND transporter family member 3	147.29
					ENSSSCG00000025667	FBXO2	F-box protein 2	38.80
					ENSSSCG00000003421	FBXO6	F-box protein 6	14.01
					ENSSSCG00000003419	MAD2L2	Mitotic arrest deficient 2 like 2	9.05
					ENSSSCG00000003423	DRAXIN	Dorsal inhibitory axon guidance protein	Within
					ENSSSCG00000022401	AGTRAP	Angiotensin II receptor associated protein	37.13
					ENSSSCG00000046656	NA	NA	57.69
					ENSSSCG00000003428	MTHFR	Methylenetetrahydrofolate reductase	75.40
					ENSSSCG00000003429	CLCN6	Chloride voltage-gated channel 6	93.74
					ENSSSCG00000003431	NPPB	Natriuretic peptide B	131.26
					ENSSSCG00000003432	KIAA2013	KIAA2013	187.31
					ENSSSCG00000028965	U5	SnRNA	182.81
MARC0058875	7	24865378	4.99E-09	−0.57	ENSSSCG00000030874	NA	NA	153.33
					ENSSSCG00000033414	NA	NA	105.83
					ENSSSCG00000027921	NA	NA	72.79
					ENSSSCG00000001447	NA	NA	75.19
					ENSSSCG00000025071	BTNL2	Butyrophilin like 2	55.52
					ENSSSCG00000001453	HLA-DRA	SLA-DRA:MHC class II DR-alpha	30.22
					ENSSSCG00000001455	SLA-DRB1	MHC class II histocompatibility antigen SLA-DRB1	16.40
					ENSSSCG00000001457	SLA-DQB1	SLA-DQ beta1 domain	77.73
					ENSSSCG00000001456	SLA-DQA1	MHC class II histocompatibility antigen SLA-DQA	86.12
					ENSSSCG00000001459	HLA-DOB	SLA-DOB:MHC class II, DO beta	165.06
					ENSSSCG00000025593	TAP2	Transporter 2, ATP binding cassette subfamily B member	178.34
					ENSSSCG00000026951	PSMB8	Proteasome 20S subunit beta 8	187.75
					ENSSSCG00000001463	PSMB9	Proteasome 20S subunit beta 9	190.22
					ENSSSCG00000025618	TAP1	Transporter 1, ATP binding cassette subfamily B member	197.40
WU_10.2_10_49571394	10	44833617	6.79E-08	0.76	ENSSSCG00000011040	CACNB2	Calcium voltage-gated channel auxiliary subunit beta 2	Within
					ENSSSCG00000011041	NSUN6	Putative methyltransferase NSUN6	48.83
					ENSSSCG00000040106	NA	NA	155.91
M1GA0014659	11	4191013	1.19E-07	0.36	ENSSSCG00000024064	RNF6	Ring finger protein 6	158.85
					ENSSSCG00000009298	CDK8	Cyclin dependent kinase 8	9.92
					ENSSSCG00000009300	WASF3	WASP family member 3	70.16
MARC0042106	15	3073986	1.16E-07	0.43	ENSSSCG00000015677	LYPD6B	LY6/PLAUR domain containing 6B	101.07
					ENSSSCG00000022919	KIF5C	Kinesin family member 5C	116.36

Note: SSC: Sus scrofa *Chromosome; NA: indicates novel gene in Ensembl database.*

performed the GWAS and proposed seven significant SNPs associated with sow milk production ability. By the estimation of least-square means, ASGA0010040, MARC0058875, WU_10.2_10_49571394, M1GA0014659, and MARC0042106 were found that the extreme phenotypic values significantly corresponded to the homozygous genotypes. Sows that were genotyped for MARC0029355 and WU_10.2_10_49571394 had an obvious phenotype trend between two different homozygous, while not significant. This might be due to the high SE.

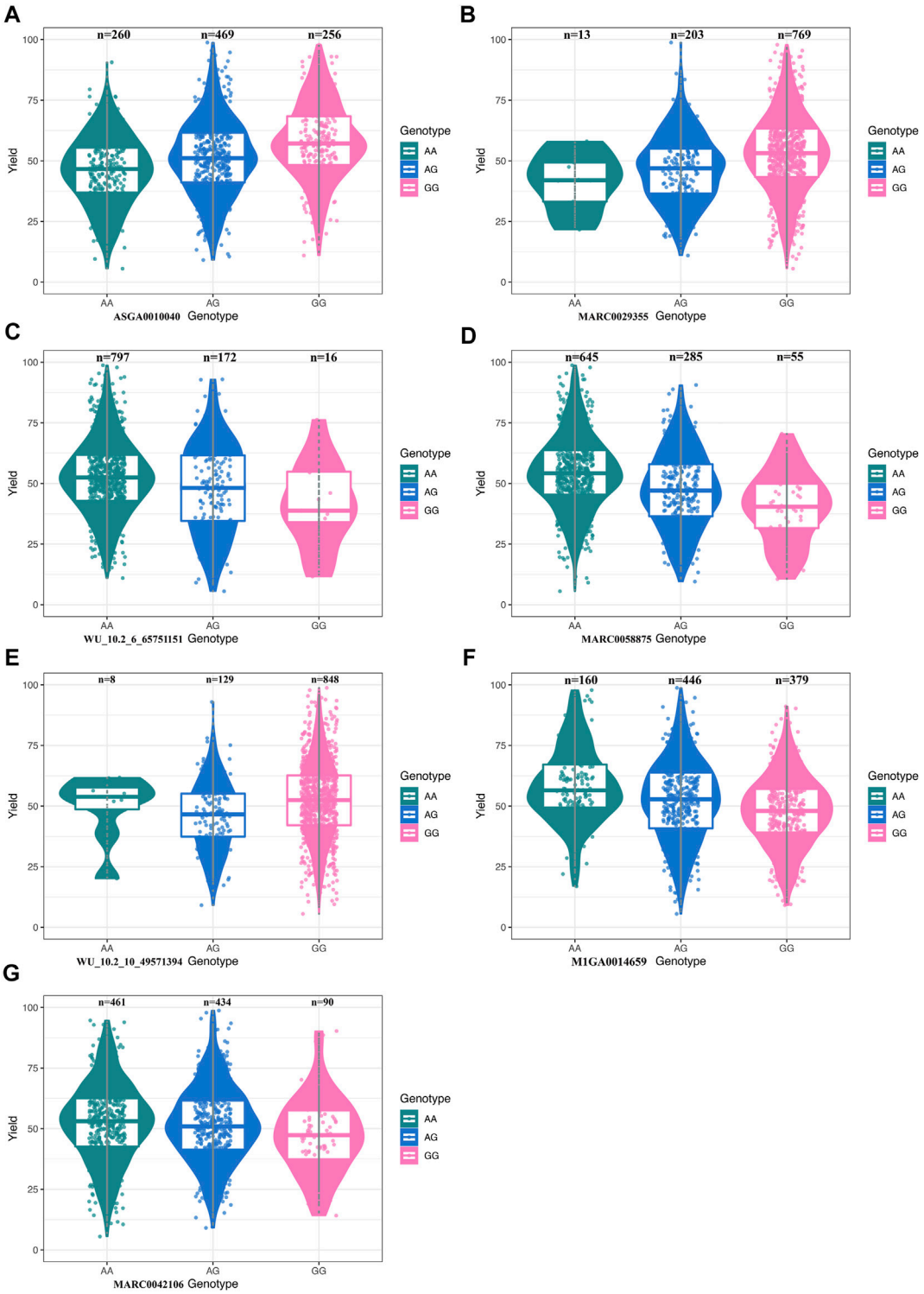


FIGURE 3 | (A) Boxplot for litter weight gain (LWG) and the genotype at SNP ASGA0010040. **(B)** Boxplot for LWG and the genotype at SNP MARC0029355. **(C)** Boxplot for LWG and the genotype at SNP WU_10.2_6_65751151. **(D)** Boxplot for LWG and the genotype at SNP MARC0058875. **(E)** Boxplot for LWG and the genotype at SNP WU_10.2_10_49571394. **(F)** Boxplot for LWG and the genotype at SNP M1GA0014659. **(G)** Boxplot for LWG and the genotype at SNP MARC0042106.

TABLE 2 | Least square mean (± SE) of sow litter weight gain (LWG) by the genotype of seven significant SNPs.

SNP	Genotypes	*NO.*	Frequency	Allele	Frequency	Sow milk production ability (kg)
ASGA0010040	AA	260	0.2640	A	0.5020	46.8144 ± 1.1143^{Aa}
	AG	469	0.4761			49.2446 ± 0.9922^{b}
	GG	256	0.2599	G	0.4980	50.7648 ± 1.1816^{Bb}
MARC0029355	AA	13	0.0132	A	0.1162	45.5502 ± 3.4420
	AG	203	0.2061			48.1851 ± 1.1887
	GG	769	0.7807	G	0.8838	49.0730 ± 0.9386
WU_10.2_6_65751151	AA	797	0.8091	A	0.8964	49.7458 ± 0.9377^{B}
	AG	172	0.1746			45.8905 ± 1.2061^{A}
	GG	16	0.0162	G	0.1036	46.1629 ± 3.1500^{AB}
MARC0058875	AA	645	0.6548	A	0.7995	50.1935 ± 0.9668^{B}
	AG	285	0.2893			47.1476 ± 1.0909^{A}
	GG	55	0.0558	G	0.2005	44.1264 ± 1.8422^{A}
WU_10.2_10_49571394	AA	8	0.0081	A	0.0736	56.2333 ± 4.3318
	AG	129	0.1310			49.9266 ± 1.3490
	GG	848	0.8609	G	0.9264	48.4391 ± 0.9337
M1GA0014659	AA	160	0.1624	A	0.3888	50.7410 ± 1.2967^{b}
	AG	446	0.4528			48.6969 ± 1.0052^{ab}
	GG	379	0.3848	G	0.6112	48.1426 ± 1.0410^{a}
MARC0042106	AA	461	0.4680	A	0.6883	47.8839 ± 1.0019^{a}
	AG	434	0.4406			49.6625 ± 1.0058^{b}
	GG	90	0.0914	G	0.3117	49.3675 ± 1.5408^{ab}

Note: No.*: Number of cows with corresponding genotypes. Different letter (small letters:* $p < 0.05$*; capital letters:* $p < 0.01$*) superscripts indicate significant differences among the genotypes.*

The lactation process includes initiation and maintenance, which are mainly regulated by hormone-nerve. Milk production is highly influenced by the sow's body reserves at the start of lactation as well as the degree and type of body tissues that are mobilized during lactation (Costermans et al., 2020). Selection for high prolificacy in modern sows has led to increased litter size and a higher number of piglets weaned per litter, which results in greater metabolic demands during lactation, due to a higher milk production (Kemp et al., 2018). In our research, we found the candidate genes were enriched mainly in metabolism-related functions, especially in processes involving carbohydrates, ATP, lipids, and protein processes. In addition, we also found that these candidate genes were involved in colostrum immune processes and milk synthesis.

By the combined analysis with the swine mammary gland transcriptome data, nine genes were identified to be key candidates. By the combined analysis with the reported GWAS data, the NAV2 gene was found to be located with 0.86 Mb of the reported QTL region ssc2: 40936355. We comprehensively analyzed the results of functional enrichments, the swine mammary gland transcriptome, and previous GWAS data, which revealed that 28 candidate genes were associated with swine milk production, and 10 of them were key candidates.

For the 10 key candidate genes, NAV2 was mainly enriched into Na (+) channel (Mishra et al., 2015), nervous system development (Clagett-Dame et al., 2006; Yan et al., 2015; Pook et al., 2020), and delayed age of menopause among women (Bae et al., 2019). In all brain regions studied, the levels of NAV2 observed in late gestation and early postnatal life were the highest (Pook et al., 2020). It was reported that NAV2 was associated with hyperlipidemia (Sun et al., 2018a). ANO3 was associated with dystonia and motor neuron dysfunction (García-Hernández et al., 2021). The glycoprotein MUC15 was initially isolated from the bovine milk fat globule membrane and had a potential physiological function in signal transduction (Pallesen et al., 2008). MUC15 was involved in PI3K/AKT signaling pathway (Yue et al., 2020), and the localization of MUC15 was shown to be controlled by the ovarian hormones, oestrogen, and progesterone (Poon et al., 2014). DISP3 was a molecule between thyroid hormone and cholesterol metabolism, which used thyroid hormone to regulate serum cholesterol levels, thus participating in the metabolism and synthesis of various substances such as sugar, protein, fat, estradiol, and cortisol in the body (Zikova et al., 2009). DISP3 was also associated with the release of lipid-anchored secretory proteins (Katoh and Katoh, 2005). FBXO6 was related to ovarian cancer treatment (Ji et al., 2021) and glycoprotein quality control (Glenn et al., 2008). CLCN6 was involved in the renin-angiotensin-aldosterone system (Ji et al., 2017). SLA-DRA, SLA-DRB1, SLA-DQB1, and SLA-DQA1 were the SLA class Ⅱ genes involved in immune (Liu et al., 2015).

TABLE 3 | Results of the combined analysis with the reported swine mammary gland transcriptome and milk production ability GWAS data.

Corresponding genes of candidate genes located in the reported QTLs						Results of the combined analysis between the previous RNA-seq and the current GWAS		
Gene	**Gene symbol**	**QTL (bp)**	**Distance (bp)**	**Distance (Mb)**	**Traits (reference)**	**Gene**	**Group**	***p*-value**
ENSSSCG00000013351	NAV2	ssc2: 40936355	858430	0.86	LWG and EOP[b]	ANO3	(−10) vs (−14)	1.00E+00
ENSSSCG00000013338	SLC5A12	ssc2: 40936355	7322806	7.32	LWG and EOP[b]		(−6) vs (−14)	1.08E-01
ENSSSCG00000013339	ANO3	ssc2: 40936355	6949145	6.95	LWG and EOP[b]		(−2) vs (−14)	9.87E-03
ENSSSCG00000013340	MUC15	ssc2: 40936355	7237662	7.24	LWG and EOP[b]		(+1) vs (−14)	4.00E-01
ENSSSCG00000030874	NA	ssc7: 94754228	70042179	70.04	LWG[a]	MUC15	(−10) vs (−14)	1.00E+00
ENSSSCG00000030874	NA	ssc7: 118733319	94021270	94.02	LWG[a]		(−6) vs (−14)	3.81E-01
ENSSSCG00000027921	NA	ssc7: 94754228	69961637	69.96	LWG[a]		(−2) vs (−14)	2.89E-02
ENSSSCG00000027921	NA	ssc7: 118733319	93940728	93.94	LWG[a]		(+1) vs (−14)	1.49E-02
ENSSSCG00000001447	NA	ssc7: 94754228	69964042	69.96	LWG[a]	DISP3	(−10) vs (−14)	1.00E+00
ENSSSCG00000001447	NA	ssc7: 118733319	93943133	93.94	LWG[a]		(−6) vs (−14)	5.06E-01
ENSSSCG00000025071	BTNL2	ssc7: 94754228	69944374	69.94	LWG[a]		(−2) vs (−14)	1.04E-01
ENSSSCG00000025071	BTNL2	ssc7: 118733319	93923465	93.92	LWG[a]		(+1) vs (−14)	1.72E-02
ENSSSCG00000001453	HLA-DRA	ssc7: 94754228	69919068	69.92	LWG[a]	FBXO6	(−10) vs (−14)	1.00E+00
ENSSSCG00000001453	HLA-DRA	ssc7: 118733319	93898159	93.90	LWG[a]		(−6) vs (−14)	4.33E-01
ENSSSCG00000001455	SLA-DRB1	ssc7: 94754228	69840175	69.84	LWG[a]		(−2) vs (−14)	6.96E-01
ENSSSCG00000001455	SLA-DRB1	ssc7: 118733319	93819266	93.82	LWG[a]		(+1) vs (−14)	5.00E-02
ENSSSCG00000001457	SLA-DQB1	ssc7: 94754228	69776931	69.78	LWG[a]	CLCN6	(−10) vs (−14)	1.00E+00
ENSSSCG00000001457	SLA-DQB1	ssc7: 118733319	93756022	93.76	LWG[a]		(−6) vs (−14)	6.93E-01
ENSSSCG00000001456	SLA-DQA1	ssc7: 94754228	69758591	69.76	LWG[a]		(−2) vs (−14)	1.67E-01
ENSSSCG00000001456	SLA-DQA1	ssc7: 118733319	93737682	93.74	LWG[a]		(+1) vs (−14)	8.98E-02
ENSSSCG00000001459	HLA-DOB	ssc7: 94754228	69716277	69.72	LWG[a]	HLA-DRA	(−10) vs (−14)	1.00E+00
ENSSSCG00000001459	HLA-DOB	ssc7: 118733319	93695368	93.70	LWG[a]		(−6) vs (−14)	8.86E-01
ENSSSCG00000025593	TAP2	ssc7: 94754228	69697080	69.70	LWG[a]		(−2) vs (−14)	1.14E-01
ENSSSCG00000025593	TAP2	ssc7: 118733319	93676171	93.68	LWG[a]		(+1) vs (−14)	1.59E-03
ENSSSCG00000026951	PSMB8	ssc7: 94754228	69681182	69.68	LWG[a]	SLA-DQB1	(−10) vs (−14)	1.00E+00
ENSSSCG00000026951	PSMB8	ssc7: 118733319	93660273	93.66	LWG[a]		(−6) vs (−14)	6.19E-01
ENSSSCG00000001463	PSMB9	ssc7: 94754228	69675634	69.68	LWG[a]		(−2) vs (−14)	2.63E-01
ENSSSCG00000001463	PSMB9	ssc7: 118733319	93654725	93.65	LWG[a]		(+1) vs (−14)	3.72E-02
ENSSSCG00000025618	TAP1	ssc7: 94754228	69682366	69.68	LWG[a]	SLA-DQA1	(−10) vs (−14)	1.00E+00
ENSSSCG00000025618	TAP1	ssc7: 118733319	93661457	93.66	LWG[a]		(−6) vs (−14)	3.27E-01
							(−2) vs (−14)	3.18E-02
							(+1) vs (−14)	2.79E-04
						SLA-DRB1	(−10) vs (−14)	1.00E+00
							(−6) vs (−14)	6.56E-01
							(−2) vs (−14)	2.66E-01
							(+1) vs (−14)	1.63E-02

Note: GWAS: Genome-wide association study. QTL: Quantitative trait loci. NA: Novel gene in Ensembl database. (−14), (−10), (−6), (−2), and (+ 1): At days 14, 10, 6, and 2 before (−) parturition and day 1 after (+) parturition. a: The reference ~ Thekkoot DM, Young JM, Rothschild MF, Dekkers JC: Genomewide association analysis of sow lactation performance traits in lines of Yorkshire pigs divergently selected for residual feed intake during grow-finish phase. J Anim Sci *2016, 94 (6):2317–2331. b: The reference ~ Chapter 4. A genome wide association analysis for sow lactation traits in Yorkshire and Landrace sows (https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=5224&context=etd). c: The reference ~ Palombo V, Loor JJ, D'Andrea M, Vailati-Riboni M, Shahzad K, Krogh U, Theil PK: Transcriptional profiling of swine mammary gland during the transition from colostrogenesis to lactogenesis using RNA sequencing.* Bmc Genomics *2018, 19.*

SLC5A12 was an active source of lactate transmembrane transporter, which is mainly involved in sodium ion transport (Martin et al., 2007; Sivaprakasam et al., 2017). FBXO2 and MAD2L2 were involved in ubiquitination processes (Li et al., 2018; Liu et al., 2021), which regulated the milk protein and fat metabolic mechanism (Liu et al., 2020a). DRAXIN was related to Akt, which could impact milk synthesis (Meli et al., 2015; Liu et al., 2020b). AGTRAP was reported to have a functional role in adipose metabolism (Ohki et al., 2017). MTHFR was involved in the metabolism of carbon, methionine, and tetrahydrofolic acid, and was related to the metabolism of milk folic acid (Page et al., 2019). MTHFR could play a role in milk protein synthesis through folic acid (Hou et al., 2015). Studies reported that MTHFR was an important candidate gene for sheep milk yield traits (Hou et al., 2015; An et al., 2016). NPPB and BTNL2 were involved in PI3K/AKT, Ca^{2+}, K^{+}, ATP, and immunity (Fioretti et al., 2004; Dolovcak et al., 2009; Sun et al., 2018b; Zhao et al., 2020). KIAA2013 was related to DNA methylation levels of newborns (Yeung et al., 2021). HLA-DOB, PSMB8, and TAP1 were involved in immune, protein and fat metabolism processes (Nagarajan et al., 2002; Niesporek et al., 2005; Garg, 2011; Kolbus et al., 2012; Arimochi et al., 2016; Naderi et al., 2016; Moussa et al., 2018; Yang et al., 2018; Chen et al., 2020). CACNB2 was involved in the regulation of ion membrane transport, which was

related to calcium channel activity, MAPK, and oxytocin signaling pathways (Durairaj Pandian et al., 2019), and studies have shown that CACNB2 was involved in the formation of porcine marlin (Bertolini et al., 2018). NSUN6 protein might have an important function in broad aspects of embryonic development (Chi and Delgado-Olguín, 2013). KIF5C was involved in the regulation of mammalian phosphorylation (Padzik et al., 2016). As the substrate of protein kinase CK2, KIF5C cloud interacts with CK2alpha to become a negative factor of adipogenesis (Schäfer et al., 2008; Chen et al., 2017).

ENSSSCG00000030874, ENSSSCG00000027921, and ENSSSCG00000001447 genes were novel genes in the Ensembl database, while our functional analysis showed their roles in the immune system.

In conclusion, we identified seven SNPs significantly associated with sow milk production ability and propose 28 candidate genes. By integrated analysis of the biological functions, swine mammary gland transcriptome, and previous GWAS data, 10 genes (NAV2, ANO3, MUC15, DISP3, FBXO6, CLCN6, HLA-DRA, SLA-DQB1, HLA-DRB1, SLA-DQB1, and SLA-DQA1) were proposed to the key candidates. Our study provided a new insight for investigating the potential critical SNPs and genes involved in sow milk production, and the molecular information might be used to improve sow lactation performance.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal experiments were approved by the Science Research Department of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS) (Beijing, China).

AUTHOR CONTRIBUTIONS

LXW and LS conceived and designed the study. LS collected the DNA and phenotype samples with the help of YL, QL, LZ, LGW, XL, HG, XH, FZ, and HY. LS analyzed the data and prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was supported by the China Agriculture Research System of MOF and MARA, and Chinese Academy of Agricultural Sciences Foundation (20118/2020-YWF-YTS-8).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.724533/full#supplementary-material

REFERENCES

An, X., Song, Y., Hou, J., Wang, S., Gao, K., and Cao, B. (2016). Identification of a Functional SNP in the 3′-UTR of CaprineMTHFRgene That Is Associated With Milk Protein Levels. *Anim. Genet.* 47, 499–503. doi:10.1111/age.12425

Arimochi, H., Sasaki, Y., Kitamura, A., and Yasutomo, K. (2016). Differentiation of Preadipocytes and Mature Adipocytes Requires PSMB8. *Sci. Rep.* 6, 26791. doi:10.1038/srep26791

Auldist, D. E., Morrish, L., Eason, P., and King, R. H. (1998). The Influence of Litter Size on Milk Production of Sows. *Anim. Sci.* 67, 333–337. doi:10.1017/s1357729800010109

Bae, H., Lunetta, K. L., Murabito, J. M., Andersen, S. L., Schupf, N., Perls, T., et al. (2019). Genetic Associations With Age of Menopause in Familial Longevity. *Menopause.* 26, 1204–1212. doi:10.1097/gme.0000000000001367

Baxter, E., Rutherford, K., D'Eath, R., Arnott, G., Turner, S., Sandøe, P., et al. (2013). The Welfare Implications of Large Litter Size in the Domestic Pig II: Management Factors. *Anim. Welfare.* 22, 219–238. doi:10.7120/09627286.22.2.219

Bergsma, R., Kanis, E., Verstegen, M. W. A., and Knol, E. F. (2008). Genetic Parameters and Predicted Selection Results for Maternal Traits Related to Lactation Efficiency in Sows. *J. Anim. Sci.* 86, 1067–1080. doi:10.2527/jas.2007-0165

Bertolini, F., Schiavo, G., Galimberti, G., Bovo, S., D'Andrea, M., Gallo, M., et al. (2018). Genome-wide Association Studies for Seven Production Traits Highlight Genomic Regions Useful to Dissect Dry-Cured Ham Quality and Production Traits in Duroc Heavy Pigs. *Animal.* 12, 1777–1784. doi:10.1017/s1751731118000757

Chen, M., Li, Y., Lv, H., Yin, P., Zhang, L., and Tang, P. (2020). Quantitative Proteomics and Reverse Engineer Analysis Identified Plasma Exosome Derived Protein Markers Related to Osteoporosis. *J. Proteomics.* 228, 103940. doi:10.1016/j.jprot.2020.103940

Chen, Q., Hao, W., Xiao, C., Wang, R., Xu, X., Lu, H., et al. (2017). SIRT6 Is Essential for Adipocyte Differentiation by Regulating Mitotic Clonal Expansion. *Cel Rep.* 18, 3155–3166. doi:10.1016/j.celrep.2017.03.006

Chi, L., and Delgado-Olguín, P. (2013). Expression of NOL1/NOP2/sun Domain (Nsun) RNA Methyltransferase Family Genes in Early Mouse Embryogenesis. *Gene Expr. Patterns.* 13, 319–327. doi:10.1016/j.gep.2013.06.003

Clagett-Dame, M., McNeill, E. M., and Muley, P. D. (2006). Role of All-Trans Retinoic Acid in Neurite Outgrowth and Axonal Elongation. *J. Neurobiol.* 66, 739–756. doi:10.1002/neu.20241

Costermans, N. G. J., Soede, N. M., Middelkoop, A., Laurenssen, B. F. A., Koopmanschap, R. E., Zak, L. J., et al. (2020). Influence of the Metabolic State During Lactation on Milk Production in Modern Sows. *Animal.* 14, 2543–2553. doi:10.1017/s1751731120001536

Dolovcak, S., Waldrop, S. L., Fitz, J. G., and Kilic, G. (2009). 5-Nitro-2-(3-Phenylpropylamino)Benzoic Acid (NPPB) Stimulates Cellular ATP Release Through Exocytosis of ATP-Enriched Vesicles. *J. Biol. Chem.* 284, 33894–33903. doi:10.1074/jbc.m109.046193

Durairaj Pandian, V., Giovannucci, D. R., Vazquez, G., and Kumarasamy, S. (2019). CACNB2 Is Associated With Aberrant RAS-MAPK Signaling in Hypertensive Dahl Salt-Sensitive Rats. *Biochem. Biophysical Res. Commun.* 513, 760–765. doi:10.1016/j.bbrc.2019.03.215

Eissen, J. J., Kanis, E., and Kemp, B. (2000). Sow Factors Affecting Voluntary Feed Intake during Lactation. *Livestock Prod. Sci.* 64, 147–165. doi:10.1016/s0301-6226(99)00153-0

Elsley, F. W. H. (1971). "Nutrition and Lactation in the Sow," in *Lactation*. Editor I. R. Faulkner (London: Butterworths), 393–411.

Falker-Gieske, C., Blaj, I., Preuss, S., Bennewitz, J., Thaller, G., and Tetens, J. (2019). GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3-Genes Genom Genet.* 9, 2823–2834. doi:10.1534/g3.119.400452

Fioretti, B., Castigli, E., Calzuola, I., Harper, A. A., Franciolini, F., and Catacuzzeno, L. (2004). NPPB Block of the Intermediate-Conductance Ca2+-Activated K+ Channel. *Eur. J. Pharmacol.* 497, 1–6. doi:10.1016/j.ejphar.2004.06.034

García-Hernández, J. L., Corchete, L. A., Marcos-Alcalde, Í., Gómez-Puertas, P., Fons, C., and Lazo, P. A. (2021). Pathogenic Convergence of CNVs in Genes Functionally Associated to a Severe Neuromotor Developmental Delay Syndrome. *Hum. Genomics.* 15, 11. doi:10.1186/s40246-021-00309-4

Garg, A. (2011). Lipodystrophies: Genetic and Acquired Body Fat Disorders. *J. Clin. Endocr. Metab.* 96, 3313–3325. doi:10.1210/jc.2011-1159

Glenn, K. A., Nelson, R. F., Wen, H. M., Mallinger, A. J., and Paulson, H. L. (2008). Diversity in Tissue Expression, Substrate Binding, and SCF Complex Formation for a Lectin Family of Ubiquitin Ligases. *J. Biol. Chem.* 283, 12717–12729. doi:10.1074/jbc.m709508200

Goddard, M. E., and Hayes, B. J. (2009). Mapping Genes for Complex Traits in Domestic Animals and Their Use in Breeding Programmes. *Nat. Rev. Genet.* 10, 381–391. doi:10.1038/nrg2575

Goldman, A. S. (2002). Evolution of the Mammary Gland Defense System and the Ontogeny of the Immune System. *J. Mammary Gland Biol. Neoplasia.* 7, 277–289. doi:10.1023/a:1022852700266

Hou, J., An, X., Song, Y., Gao, T., Lei, Y., and Cao, B. (2015). Two Mutations in the Caprine MTHFR 3'UTR Regulated by MicroRNAs Are Associated With Milk Production Traits. *PLoS One.* 10, e0133015. doi:10.1371/journal.pone.0133015

Ji, L.-D., Li, J.-Y., Yao, B.-B., Cai, X.-B., Shen, Q.-J., and Xu, J. (2017). Are Genetic Polymorphisms in the Renin-Angiotensin-Aldosterone System Associated With Essential Hypertension? Evidence From Genome-Wide Association Studies. *J. Hum. Hypertens.* 31, 695–698. doi:10.1038/jhh.2017.29

Ji, M., Zhao, Z., Li, Y., Xu, P., Shi, J., Li, Z., et al. (2021). FBXO6-mediated RNASET2 Ubiquitination and Degradation Governs the Development of Ovarian Cancer. *Cell Death Dis.* 12, 317. doi:10.1038/s41419-021-03580-4

Katoh, Y., and Katoh, M. (2005). Identification and Characterization of DISP3 Gene In Silico. *Int. J. Oncol.* 26, 551–556. doi:10.3892/ijo.26.2.551

Kemp, B., Da Silva, C. L. A., and Soede, N. M. (2018). Recent Advances in Pig Reproduction: Focus on Impact of Genetic Selection for Female Fertility. *Reprod. Dom Anim.* 53 (Suppl. 2), 28–36. doi:10.1111/rda.13264

Kolbus, D., Ljungcrantz, I., Söderberg, I., Alm, R., Björkbacka, H., Nilsson, J., et al. (2012). TAP1-Deficiency Does Not Alter Atherosclerosis Development in Apoe−/− Mice. *PLoS One.* 7, e33932. doi:10.1371/journal.pone.0033932

Li, Y., Li, L., Chen, M., Yu, X., Gu, Z., Qiu, H., et al. (2018). MAD2L2 Inhibits Colorectal Cancer Growth by Promoting NCOA3 Ubiquitination and Degradation. *Mol. Oncol.* 12, 391–405. doi:10.1002/1878-0261.12173

Liu, L. L., Guo, A. W., Li, Q., Wu, P. F., Yang, Y., Chen, F. F., et al. (2020a). The Regulation of Ubiquitination in Milk Fat Synthesis in Bovine. *Yi Chuan.* 42, 548–555. doi:10.16288/j.yczz.20-037

Liu, Y., Hou, J., Zhang, M., Seleh-Zo, E., Wang, J., Cao, B., et al. (2020b). circ-016910 Sponges miR-574-5p to Regulate Cell Physiology and Milk Synthesis via MAPK and PI3K/AKT-mTOR Pathways in GMECs. *J. Cel Physiol* 235, 4198–4216. doi:10.1002/jcp.29370

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *Plos Genet.* 12, e1005767. doi:10.1371/journal.pgen.1005767

Liu, Y., Pan, B., Qu, W., Cao, Y., Li, J., and Zhao, H. (2021). Systematic Analysis of the Expression and Prognosis Relevance of FBXO Family Reveals the Significance of FBXO1 in Human Breast Cancer. *Cancer Cel Int.* 21, 130. doi:10.1186/s12935-021-01833-y

Liu, Z. Z., Xia, J. H., Xin, L. L., Wang, Z. G., Qian, L., Wu, S. G., et al. (2015). Swine Leukocyte Antigen Class II Genes (SLA-DRA, SLA-DRB1, SLA-DQA, SLA-DQB1) Polymorphism and Genotyping in Guizhou Minipigs. *Genet. Mol. Res.* 14, 15256–15266. doi:10.4238/2015.november.30.1

Martin, P. M., Dun, Y., Mysona, B., Ananth, S., Roon, P., Smith, S. B., et al. (2007). Expression of the Sodium-Coupled Monocarboxylate Transporters SMCT1 (SLC5A8) and SMCT2 (SLC5A12) in Retina. *Invest. Ophthalmol. Vis. Sci.* 48, 3356–3363. doi:10.1167/iovs.06-0888

Meli, R., Weisová, P., and Propst, F. (2015). Repulsive Axon Guidance by Draxin Is Mediated by Protein Kinase B (Akt), Glycogen Synthase Kinase-3β (GSK-3β) and Microtubule-Associated Protein 1B. *PLoS One.* 10, e0119524. doi:10.1371/journal.pone.0119524

Mishra, S., Reznikov, V., Maltsev, V. A., Undrovinas, N. A., Sabbah, H. N., and Undrovinas, A. (2015). Contribution of Sodium Channel Neuronal Isoform Nav1.1 to Late Sodium Current in Ventricular Myocytes From Failing Hearts. *J. Physiol.* 593, 1409–1427. doi:10.1113/jphysiol.2014.278259

Moussa, E. M., Huang, H., Thézénas, M. L., Fischer, R., Ramaprasad, A., Sisay-Joof, F., et al. (2018). Proteomic Profiling of the Plasma of Gambian Children With Cerebral Malaria. *Malar. J.* 17, 337. doi:10.1186/s12936-018-2487-y

Naderi, M., Hashemi, M., and Aminini, S. (2016). Association of TAP1 and TAP2 Gene Polymorphisms with Susceptibility to Pulmonary Tuberculosis. *Iran J. Allergy Asthma Immunol.* 15, 62–68.

Nagarajan, U. M., Lochamy, J., Chen, X., Beresford, G. W., Nilsen, R., Jensen, P. E., et al. (2002). Class II Transactivator Is Required for Maximal Expression of HLA-DOB in B Cells. *J. Immunol.* 168, 1780–1786. doi:10.4049/jimmunol.168.4.1780

Niesporek, S., Meyer, C. G., Kremsner, P. G., and May, J. (2005). Polymorphisms of Transporter Associated with Antigen Processing Type 1 (TAP1), Proteasome Subunit Beta Type 9 (PSMB9) and Their Common Promoter in African Children with Different Manifestations of Malaria. *Int. J. Immunogenet.* 32, 7–11. doi:10.1111/j.1744-313x.2005.00484.x

Ohki, K., Wakui, H., Azushima, K., Uneda, K., Haku, S., Kobayashi, R., et al. (2017). ATRAP Expression in Brown Adipose Tissue Does Not Influence the Development of Diet-Induced Metabolic Disorders in Mice. *Int. J. Mol. Sci.* 18, 676. doi:10.3390/ijms18030676

Padzik, A., Deshpande, P., Hollos, P., Franker, M., Rannikko, E. H., Cai, D., et al. (2016). KIF5C S176 Phosphorylation Regulates Microtubule Binding and Transport Efficiency in Mammalian Neurons. *Front. Cel. Neurosci.* 10, 57. doi:10.3389/fncel.2016.00057

Page, R., Wong, A., Arbuckle, T. E., and MacFarlane, A. J. (2019). The MTHFR 677C>T Polymorphism Is Associated With Unmetabolized Folic Acid in Breast Milk in a Cohort of Canadian Women. *Am. J. Clin. Nutr.* 110, 401–409. doi:10.1093/ajcn/nqz056

Pallesen, L. T., Pedersen, L. R. L., Petersen, T. E., Knudsen, C. R., and Rasmussen, J. T. (2008). Characterization of Human Mucin (MUC15) and Identification of Ovine and Caprine Orthologs. *J. Dairy Sci.* 91, 4477–4483. doi:10.3168/jds.2008-1204

Palombo, V., Loor, J. J., D'Andrea, M., Vailati-Riboni, M., Shahzad, K., Krogh, U., et al. (2018). Transcriptional Profiling of Swine Mammary Gland During the Transition from Colostrogenesis to Lactogenesis Using RNA Sequencing. *Bmc Genomics.* 19, 322. doi:10.1186/s12864-018-4719-5

Pettigrew, J. E., Cornelius, S. G., Moser, R. L., and Sower, A. F. (1987). A Refinement and Evaluation of the Isotope Dilution Method for Estimating Milk Intake by Piglets. *Livestock Prod. Sci.* 16, 163–174. doi:10.1016/0301-6226(87)90017-0

Pook, C., Ahrens, J. M., and Clagett-Dame, M. (2020). Expression Pattern of Nav2 in the Murine CNS With Development. *Gene Expr. Patterns.* 35, 119099. doi:10.1016/j.gep.2020.119099

Poon, C. E., Lecce, L., Day, M. L., and Murphy, C. R. (2014). Mucin 15 Is Lost but Mucin 13 Remains in Uterine Luminal Epithelial Cells and the Blastocyst at the Time of Implantation in the Rat. *Reprod. Fertil. Dev.* 26, 421–431. doi:10.1071/rd12313

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, 559–575. doi:10.1086/519795

Revell, D. K., Williams, I. H., Mullan, B. P., Ranford, J. L., and Smits, R. J. (1998). Body Composition at Farrowing and Nutrition during Lactation Affect the Performance of Primiparous Sows: II. Milk Composition, Milk Yield, and Pig Growth. *J. Anim. Sci.* 76, 1738–1743. doi:10.2527/1998.7671738x

Schäfer, B., Götz, C., and Montenarh, M. (2008). The Kinesin I Family Member KIF5C Is a Novel Substrate for Protein Kinase CK2. *Biochem. Biophysical Res. Commun.* 375, 179–183. doi:10.1016/j.bbrc.2008.07.107

Sivaprakasam, S., Bhutia, Y. D., Yang, S., and Ganapathy, V. (2017). Short-Chain Fatty Acid Transporters: Role in Colonic Homeostasis. *Compr. Physiol.* 8, 299–314. doi:10.1002/cphy.c170014

Spötter, A., and Distl, O. (2006). Genetic Approaches to the Improvement of Fertility Traits in the Pig. *Vet. J.* 172, 234–247. doi:10.1016/j.tvjl.2005.11.013

Sun, L., Dong, Y., Zhao, J., Yin, Y., Tong, B., Zheng, Y., et al. (2018a). NPPB Modulates Apoptosis, Proliferation, Migration and Extracellular Matrix Synthesis of Conjunctival Fibroblasts by Inhibiting PI3K/AKT Signaling. *Int. J. Mol. Med.* 41, 1331–1338. doi:10.3892/ijmm.2017.3323

Sun, R., Weng, H., Men, R., Xia, X., Chong, K. C., Wu, W. K. K., et al. (2018b). Gene-methylation Epistatic Analyses via the W-Test Identifies Enriched Signals of Neuronal Genes in Patients Undergoing Lipid-Control Treatment. *BMC Proc.* 12, 53. doi:10.1186/s12919-018-0143-8

Thekkoot, D. M., Kemp, R. A., Rothschild, M. F., Plastow, G. S., and Dekkers, J. C. M. (2016). Estimation of Genetic Parameters for Traits Associated with Reproduction, Lactation, and Efficiency in Sows1. *J. Anim. Sci.* 94, 4516–4529. doi:10.2527/jas.2015-0255

Thekkoot, D. M., Young, J. M., Rothschild, M. F., and Dekkers, J. C. M. (2016). Genomewide Association Analysis of Sow Lactation Performance Traits in Lines of Yorkshire Pigs Divergently Selected for Residual Feed Intake During Grow-Finish Phase1. *J. Anim. Sci.* 94, 2317–2331. doi:10.2527/jas.2015-0258

Wang, Y., Ding, X., Tan, Z., Ning, C., Xing, K., Yang, T., et al. (2017). Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval. *Front. Genet.* 8, 194. doi:10.3389/fgene.2017.00194

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a Web Server for Annotation and Identification of Enriched Pathways and Diseases. *Nucleic Acids Res.* 39, W316–W322. doi:10.1093/nar/gkr483

Yan, J., Kim, S., Nho, K., Chen, R., Risacher, S. L., Moore, J. H., et al. (2015). Hippocampal Transcriptome-Guided Genetic Analysis of Correlated Episodic Memory Phenotypes in Alzheimer's Disease. *Front. Genet.* 6, 117. doi:10.3389/fgene.2015.00117

Yang, B.-y., Song, J.-w., Sun, H.-z., Xing, J.-c., Yang, Z.-h., Wei, C.-y., et al. (2018). PSMB8 Regulates Glioma Cell Migration, Proliferation, and Apoptosis through Modulating ERK1/2 and PI3K/AKT Signaling Pathways. *Biomed. Pharmacother.* 100, 205–212. doi:10.1016/j.biopha.2018.01.170

Yeung, E. H., Mendola, P., Sundaram, R., Zeng, X., Guan, W., Tsai, M. Y., et al. (2021). Conception by Fertility Treatment and Offspring Deoxyribonucleic Acid Methylation. *Fertil. Sterility.* 116, 493–504. doi:10.1016/j.fertnstert.2021.03.011

Yue, Y., Hui, K., Wu, S., Zhang, M., Que, T., Gu, Y., et al. (2020). MUC15 Inhibits Cancer Metastasis via PI3K/AKT Signaling in Renal Cell Carcinoma. *Cel Death Dis.* 11, 336. doi:10.1038/s41419-020-2518-9

Zhang, Y., Zhang, J., Gong, H., Cui, L., Zhang, W., Ma, J., et al. (2019). Genetic Correlation of Fatty Acid Composition With Growth, Carcass, Fat Deposition and Meat Quality Traits Based on GWAS Data in Six Pig Populations. *Meat Sci.* 150, 47–55. doi:10.1016/j.meatsci.2018.12.008

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide Association Mapping Reveals a Rich Genetic Architecture of Complex Traits in Oryza Sativa. *Nat. Commun.* 2, 467. doi:10.1038/ncomms1467

Zhao, Y., Zheng, Q., and Jin, L. (2020). The Role of B7 Family Molecules in Maternal-Fetal Immunity. *Front. Immunol.* 11, 458. doi:10.3389/fimmu.2020.00458

Zikova, M., Corlett, A., Bendova, Z., Pajer, P., and Bartunek, P. (2009). DISP3, a Sterol-Sensing Domain-Containing Protein That Links Thyroid Hormone Action and Cholesterol Metabolism. *Mol. Endocrinol.* 23, 520–528. doi:10.1210/me.2008-0271

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 in Veterinary Science

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fvets.2021.736996

Characterization of Reproductive Microbiota of Primiparous Cows During Early Postpartum Periods in the Presence and Absence of Endometritis

Hayami Kudo[1,2], *Tomochika Sugiura*[3], *Seiya Higashi*[2], *Kentaro Oka*[2], *Motomichi Takahashi*[2], *Shigeru Kamiya*[2], *Yutaka Tamura*[1] *and Masaru Usui*[1]*

[1] Department of Health and Environmental Sciences, School of Veterinary Medicine, Rakuno Gakuen University, Hokkaido, Japan, [2] Research Department, R&D Division, Miyarisan Pharmaceutical Co., Ltd., Saitama, Japan, [3] Department of Large Animal Clinical Science, School of Veterinary Medicine, Rakuno Gakuen University, Hokkaido, Japan

OPEN ACCESS

Edited by:
Hasan Riaz,
COMSATS Institute of Information Technology, Pakistan

Reviewed by:
Ricardo Zanella,
The University of Passo Fundo, Brazil
Tara G. McDaneld,
Agricultural Research Service, United States Department of Agriculture (USDA), United States

***Correspondence:**
Masaru Usui
usuima@rakuno.ac.jp

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

Received: 09 July 2021
Accepted: 21 September 2021
Published: 18 October 2021

Citation:
Kudo H, Sugiura T, Higashi S, Oka K, Takahashi M, Kamiya S, Tamura Y and Usui M (2021) Characterization of Reproductive Microbiota of Primiparous Cows During Early Postpartum Periods in the Presence and Absence of Endometritis. Front. Vet. Sci. 8:736996. doi: 10.3389/fvets.2021.736996

Endometritis has a major impact on fertility in postpartum dairy cows. Since previous studies showed an association between reproductive microbiota and perinatal disease, we monitored both bovine uterine and vaginal microbiota in primiparous cows to elucidate the effect of early postpartum microbiota on endometritis. Uterine and vaginal samples were collected at time points from pre-calving to 35 days postpartum (DPP), and analyzed by 16S rRNA sequencing, combined with ancillary bacterial culture. A total of seven healthy cows and seven cows diagnosed with endometritis on 35 DPP were used in the current study. The uterine and vaginal microbiota showed a maximum of 20.1% shared amplicon sequence variants (ASVs) at linked time points. 16S rRNA based analysis and traditional culture methods revealed that *Trueperella* showed a higher abundance in both uterus and vagina of the endometritis group compared to the healthy group on 21 DPP (U-test $p < 0.05$). Differential abundance analysis of the uterine microbiota showed that *Enterococcus* and six bacterial genera including *Bifidobacterium* were unique to the healthy group on the day of calving (0 DPP) and 28 DPP, respectively. In contrast, *Histophilus* and *Mogibacteriaceae* were characteristic bacteria in the vagina pre-calving in cows that later developed endometritis, suggesting that these bacteria could be valuable to predict clinical outcomes. Comparing the abundances of bacterial genera in the uterine microbiota, a negative correlation was observed between *Trueperella* and several bacteria including *Lactobacillus*. These results suggest that building an environment where there is an increase in bacteria that are generally recognized as beneficial, such as *Lactobacillus*, may be one possible solution to reduce the abundance of *Trueperella* and control endometritis.

Keywords: endometritis, reproductive tract microbiota, *Trueperella pyogenes*, early postpartum periods, probiotics

INTRODUCTION

Endometrial inflammation is common in postpartum dairy cows (1). Persistent uterine inflammation is clinically defined as endometritis. This modulates ovarian function and has a negative impact on fertility, resulting in high economic losses (2). The gold standard for its diagnosis after 21 days postpartum (DPP), is palpation and scoring by vaginal mucus and polymorphonuclear leukocyte (PMN) count (3, 4). Antimicrobial agents including disinfectants are routinely used in the treatment of endometritis. However, emergence and spread of antimicrobial resistance (AMR) has become a global concern for both human and livestock animal health, and appropriate use of antimicrobial agents has been strongly advocated in the veterinary field following WHO 2015 Global Action Plan on AMR, and new methods of prevention or treatment of endometritis are required (5, 6).

Recently, it has been shown that the uterus is not sterile during pregnancy and has its own microbiome (7), and non-commensal bacteria from the environment are thought to rapidly colonize the uterus shortly after calving (8). Since bacteria can be isolated from the uterus of approximately 80% of cows within 21 DPP, previous studies have focused on the late postpartum period, the time after which endometritis is established (9). After 21 DPP, *Escherichia coli*, *Trueperella pyogenes*, and *Fusobacterium necrophorum* are frequently isolated from cows with endometritis (4, 10, 11). These bacteria may interact with each other and contribute to establish the complex pathological processes of endometritis.

The uterine microbiota of the bovine postpartum period has previously been explored by culture-dependent methods. However, it is difficult to isolate bacteria with slow growth or low bacterial counts, and optimization of culture conditions is needed (12–14). With the recent progression of sequencing technology, temporal changes of uterine and vaginal microbiota in the postpartum period have been analyzed (15–18). These studies suggested that microbiota do impact on reproductive disease, although the studies did not differentiate primiparous and multiparous cows. The uterine and vaginal microbiota of multiparous cows are thought to be affected by previous parturitions (19) and therefore, studies for primiparous cows that have never experienced calving and subsequent bacterial contamination are required. In this study, to further clarify the relationship between early postpartum genital tract microbiota and endometritis, the uterine and vaginal microbiota of primiparous cows with or without endometritis were compared by sampling over the time period from pre-calving to 35 DPP.

MATERIALS AND METHODS

Animals and Diagnosis of Endometritis

This study was conducted according to the institutional guidelines for animal experiments of Rakuno Gakuen University (approval no. VH16C7). Seventeen primiparous cows (Holstein Friesian, range of age at first calving 22–27 months), received artificial insemination (AI) or embryo transfer (ET), were enrolled from December 2017 to August 2018 at the Rakuno Gakuen University for this study. All cows were diagnosed as either healthy or suffering from endometritis at 35 DPP, according to the methods previously reported (4). Briefly, clinical endometritis was diagnosed with vaginal discharge scoring as follows: 0 = clear or translucent mucus; 1 = mucus containing flecks of white or off-white pus; 2 = exudate containing <50% white or off-white mucopurulent material; and 3 = exudate containing ≥50% purulent material, usually white or yellow, occasionally bloody. The degree of recovery of the uterus was also assessed by rectal examination as follows: 0 = the diameter of largest uterine horn ≤3.5 cm and cervical diameter ≤4.5 cm; 1 = the diameter of largest uterine horn more than 3.5 to <5.5 cm and cervical diameter more than 4.5 to <7.0 cm; 2 = the diameter of largest uterine horn ≥5.5 cm and cervical diameter ≥7.0 cm. The numbers of epithelial endometrial cells and PMNs were counted to assess subclinical endometritis (3). Endometritis was defined either as total gynecological examination scoring >1, or a PMN ratio >10%. Three cows were excluded due to the following reasons: treatment with antibiotics, abnormal deliveries, and urovagina. The criteria for endometritis, scores and sampling dates of the 14 cows (seven healthy cows and seven cows diagnosed with endometritis at 35 DPP) are summarized in **Supplementary Figure 1** and **Supplementary Table 1**. In addition, parturitions were scored for difficulty and obstetric assistance required as follows: 1 = no assistance ($n = 6$); 2 = slight assistance with only one person ($n = 5$); 3 = moderate assistance with 2–3 persons ($n = 1$); 4 = severe dystocia with veterinary treatment ($n = 2$); 5 = cesarean section ($n = 0$).

Sample Collection

Both uterine and vaginal samples were collected at the same time as follows; Pre (pre-calving, only for vaginal samples), 0 DPP (within 12 h after calving and after expulsion of the placenta), 7, 21, 28, and 35 DPP (**Supplementary Table 1**).

The perineum and vulva were cleaned with a 70% ethanol swab and wiped with a paper towel. Uterine samples were collected using the cytobrush technique (3). The cytobrush device (Metribrush; Fujihira Kogyo Inc., Tokyo, Japan), double-guarded with sterilized plastic sleeve and tube, was inserted through the vagina and reached the cervix guided by palpation per rectum. Then the plastic sleeve was pulled back and the brush was moved forward and rolled over the endometrium of the uterus body. The cytobrush was retracted into the tube in uterus, then retrieved through the vagina. The tip of the cytobrush was then cut using sterile scissors, placed in 5 ml sterile saline solution and vortexed vigorously. Vaginal samples were obtained by washing methods as previously described (20), as sampling by brushing of the vaginal fornix was not suitable to extract sufficient bacterial DNA. In brief, 50 ml of sterile saline was injected into the vagina, then flushed back 2 or 3 times. Within 1 h, vaginal washings were centrifuged at 8,000 × g for 15 min and pellets were re-suspended in 5 ml sterile saline. After vortexing, 500 μl of uterine and vaginal samples were used for the culture-based method and the remaining samples were mixed in an equal volume of sterile saline containing 40% glycerol. Mixtures were flash frozen in liquid nitrogen, and stored at −80°C until use.

DNA Extraction

Bacterial DNA was extracted as described previously (21). After thawing, 2 ml of uterine and vaginal samples were mixed with 20 ml of sterile saline solution and centrifuged at 8,000 g for 15 min. Pellets were then suspended in 800 μl 10 mM Tris-HCl and 10 mM EDTA buffer containing lysozyme (Sigma-Aldrich Co., LCC, Missouri, USA) (Final concentration: 15 mg/ml). Mixtures were incubated at 37°C for 1 h. Purified achromopeptidase (Wako Pure Chemical Inc., Osaka, Japan) (Final concentration: 2,000 U/ml) was added and incubated 37°C for 30 min. Finally, proteinase K (Takara-Bio Inc., Shiga, Japan) (Final concentration: 1 mg/ml) with 20% sodium dodecyl sulfate (Sigma-Aldrich Co., LCC) was added and incubated at 55°C for 1 h. The DNA was isolated with phenol:chloroform:isoamyl alcohol (25:24:1, v/v), washed twice with 75% ethanol, and dissolved in 100 μl TE buffer. RNase A (Nippon Gene Co., Ltd., Tokyo, Japan) (Final concentration: 0.1 mg/ml) was added to the mixture, and incubated at 37°C for 1 h. Subsequently, the DNA was purified using a High Pure PCR Template Kit (Roche Inc., Basel, Switzerland) according to the manufacturer's instructions. Elution was performed in 50 μl of TE buffer and then samples were stored at −20°C until further analysis.

16S rRNA Gene Amplification and Sequencing

Extracted DNA was amplified, targeting the V3–V4 region of the bacterial 16S rRNA gene. The specific universal primer pair 341F (5′-CCTACGGGNGGCWGCAG) and 805R (5′-GACTACHVGGGTATCTAATCC) was used in this study (22). This primer set included the Illumina MiSeq sequencing adapter (forward primer: AATGATACGGCGACCACCGAGATCTACAC; reverse primer: CAAGCAGAAGACGGCATACGAGAT) and a unique barcode sequence that allowed the samples to be pooled for Illumina MiSeq sequencing. Polymerase chain reaction (PCR) was performed using MightyAmp DNA Polymerase (Takara-Bio Inc.) for 35 cycles. The PCR products were quality-checked on a 2% agarose gel electrophoresis and subsequently purified using SPRIselect beads (Beckman-Coulter Inc., California, USA). The amplified DNA was quantified using a ONEdsDNA System (Promega, Madison, WI, USA) and Quantus fluorometer (Promega). The PCR amplicon libraries were prepared by pooling approximately equal amounts of amplified DNA and sequenced on an Illumina MiSeq platform (Illumina, San Diego, USA), using the 2 × 300, v3 600-cycle kit (Illumina).

Data Processing and Analysis

Illumina Miseq fastq raw reads were analyzed with QIIME 2 platform version: 2020.2 with default scripts (23). Sequences were demultiplexed and processed by the DADA2 program (24). Before merging, reads were trimmed according to quality threshold and adapter length (Forward: 280 bp, Reverse: 220 bp). After quality filtering steps (DADA2), total, average, minimum and maximum number of non-chimeric reads were 1,901,611, 14,086, 7,565, and 17,893 reads, respectively. Amplicon sequence variants (ASVs) obtained after DADA2 [more accurate than traditional operational taxonomic units (OTUs)] were assigned using the *qiime* feature-classifier classify-sklearn method (25) against the Greengenes database (version 13.8) with 99% similarity. For diversity analysis, the number of sequence reads was rarefied to the minimum sample reads (7,565 reads) using the *qiime* diversity core-metrics-phylogenetic method (23). The rarefaction curve confirmed that the sub-sampling was sufficient to detect the bacteria species in all samples. The alpha-diversities were calculated using the Simpson and Shannon index. Beta-diversity analysis, represented by principal coordinate analysis (PCoA), was applied to the resulting weighted distance matrices to generate two-dimensional plots. For differential abundance analysis between healthy and endometritis groups, random subsampling of sequence reads was not performed.

Statistical differences in alpha-diversities between healthy and endometritis groups at sampling time points were tested using the Mann-Whitney *U*-test. The significance of the groups in the community structure was tested using permutational multivariate analysis of variance (PERMANOVA). Relative abundances of genera between healthy and endometritis groups was tested using Mann-Whitney *U*-test. To identify unique microbial genera for both healthy and endometritis groups, ANOVA-Like Differential Expression 2 (ALDEx2), estimates of the composition of biological features from the number of reads transformed to central log ratio (clr), was performed (26, 27). The correlations between the abundances of genera in uterine and vaginal microbiota were determined using Spearman's rank correlation coefficient. A linear discriminant analysis (LDA) effect size (LEfSe) (28) approach was used to identify bacterial taxonomy that was significantly different between the "before birth" group (Pre) and "after birth" group (0 DPP) for both healthy and endometritic cows.

Culture-Based Analysis

All samples used for DNA extraction were also cultured under aerobic and anaerobic conditions within 1 h from sampling. Tryptic soy (TS) agar (BD Bacto, New Jersey, USA) and blood liver (BL) agar (Nissui, Tokyo, Japan), which are widely used culture media in the analysis of human fecal microbiota (29) were adopted in this study. Both TS agar and BL agar were supplemented with 5% defibrinated horse blood. Briefly, for aerobic culture, 50 μl of 10-fold serial dilutions of sample suspensions were spread on TS agar and incubated at 37°C for 48 h. The same suspensions were spread on BL agar and incubated at 37°C for 48 h under anaerobic conditions (10% H_2, 10% CO_2, and 80% N_2) using an anaerobic chamber. After incubation, the numbers of colonies with the same morphology on agar were counted. For several different types of colonies, a maximum of three colonies with the same morphology on one plate were picked from each agar and identified using MALDI-TOF MS (Bruker, Bremen, Germany). To arbitrate the results of MALDI-TOF MS and identify bacteria to species level, 16S rRNA gene sequencing was performed by using the universal primer pair 27F (5′-AGAGTTTGATCCTGGCTCAG) and 1492R (5′-GGTTACCTTGTTACGACTT) if bacterial strains obtained an identification scoring of <2.0 (30). The obtained sequences were compared using the basic local alignment search tool (BLAST) with the National Center for Biotechnology Information (NCBI)

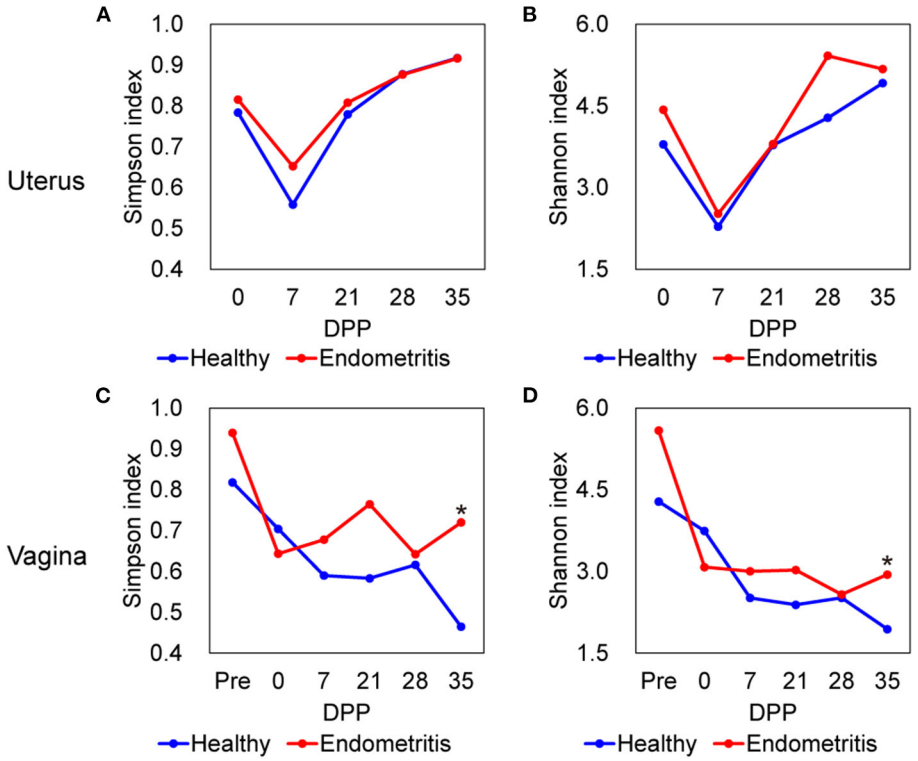


FIGURE 1 | Alpha diversity of uterine **(A,B)** and vaginal **(C,D)** microbiota in healthy (blue) and endometritis (red) groups over 35 days postpartum (DPP). **(A,C)**, Simpson index; **(B,D)**, Shannon index; * indicates significant difference at $p < 0.05$ by Mann-Whitney *U*-test.

databases. A species result was adopted only when all three colonies had the same identification result. Mann-Whitney *U*-test was used to calculate significant differences in the number of bacteria, converted to log CFU/ml of samples, between healthy and endometritis groups.

RESULTS

Alpha and Beta Diversities

The alpha-diversities of both uterine and vaginal microbiota of the healthy and endometritis groups were measured using Simpson and Shannon indices (**Figure 1**). As shown in **Figure 1**, Simpson and Shannon indices of vaginal microbiota at 35 DPP were found to be significantly higher in the endometritis group than in the healthy group (Mann-Whitney *U*-test, $p < 0.05$). As a common trend within the healthy and endometritis groups, vaginal diversity was the highest before calving and tended to decrease as time passed (Mann-Whitney *U*-test, $p < 0.01$; Pre vs. 35 DPP), while uterine diversity was lowest at 7 DPP and recovered as time passed (Mann-Whitney *U*-test, $p < 0.05$; 7 vs. 0 DPP and 7 vs. 35 DPP).

Principal coordinate analysis based on weighted UniFrac distance showed that the uterine microbiota of the healthy group was clustered and separated from those of the endometritis group at 28 and 35 DPP (PERMANOVA, $p < 0.05$) (**Supplementary Figure 2**).

Bacterial Communities in the Uterus and Vagina

A total of 3,918 ASVs were observed in samples. The ASVs shared between the uterine and vaginal microbiota of all animals at each time point are presented in **Figure 2**. Only 6.6–17.9 and 9.0–20.1% of all ASVs were shared between uterine and vaginal samples for the healthy group and endometritis group, respectively. In both the healthy and endometritis groups, the number of ASVs was higher in the vagina than in the uterus for 7 DPP. This trend was reversed after 21 DPP, namely the number of ASVs in the uterine samples was higher than that in the vaginal samples. In the healthy group, the number of ASVs at 0 DPP was almost the same in the uterine samples (207 ASVs, 34.4%) and vaginal samples (287 ASVs, 47.7%), whereas the number of ASVs in the uterine samples (414 ASVs, 60.8%) was higher than that in the vaginal samples (173 ASVs, 25.4%) in the endometritis group.

Taxonomic Composition of Uterine and Vaginal Bacterial Communities

The mean relative abundances of the top 10 bacterial genera in the uterus is displayed by bar chart (**Figure 3**). The results showed that uterine bacterial communities in the healthy group during the study period were dominated by genus *Bifidobacterium* (The relative abundance of genus *Bifidobacterium* at 0, 7, 21, and 35 DPP were 19.5, 11.5, 20.0, 31.4, and 36.4%, respectively) followed by *Streptococcus* (The relative abundance of genus *Streptococcus*

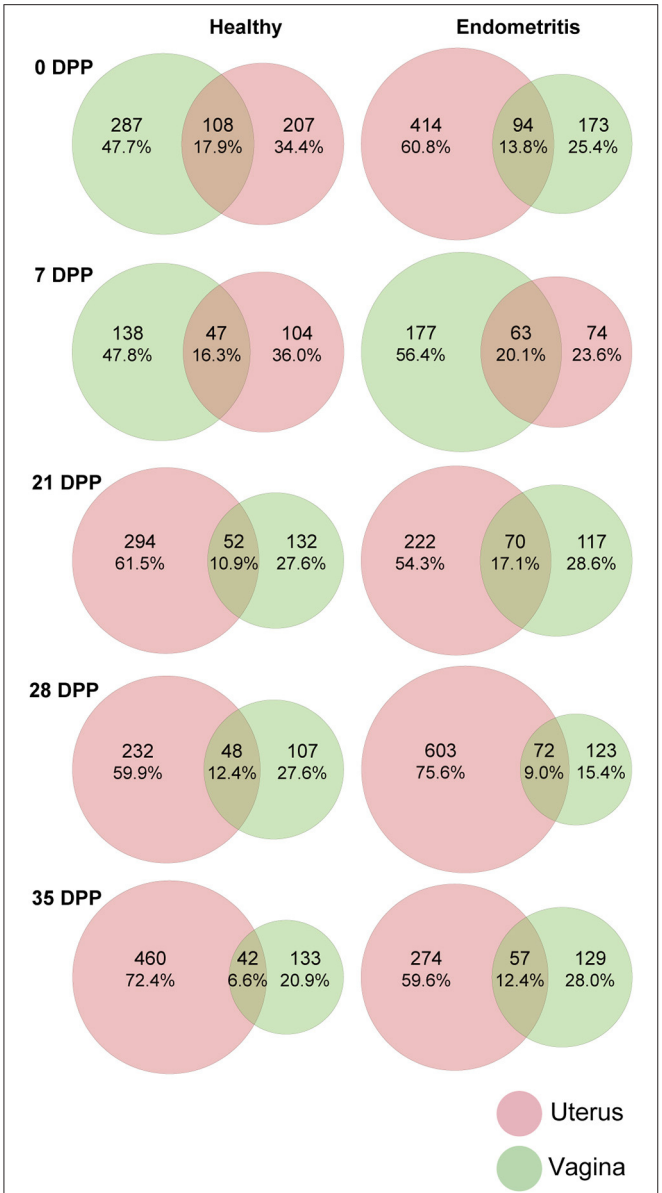


FIGURE 2 | Venn diagram of amplicon sequence variants (ASVs) of uterine (red) and vaginal (green) microbiota over 35 days postpartum (DPP). Number of ASVs in each compartment and shared ASVs between uterine and vaginal microbiota are shown after removal of <0.1% of all observed leads.

at 0, 7, 21, and 35 DPP were 17.0, 17.8, 10.3, 13.1, and 19.5%, respectively). In contrast, genus *Clostridium* was frequently detected and dominated at 0 DPP in the endometritis group (25.9%). The genus *Trueperella* was detected in both healthy and endometritis cows. With the exception that the genus *Ureaplasma* was most frequently predominant in the healthy group, vaginal bacterial communities exhibited the same dynamics as uterine in terms of genus *Bifidobacterium*, *Trueperella*, and *Clostridium* (**Supplementary Figure 3**).

These bacterial dynamics of the uterine and vaginal microbiota were also confirmed by traditional culture-dependent methods. A total of 1,213 colonies were isolated from 14 cows (healthy cows and cows with endometritis), representing 37 bacterial genera. For the three bacterial genera with the highest detection rates, bacteria count data were calculated from the number of colonies diluted 10-fold and converted to log CFU/ml of samples (**Supplementary Table 2**). As shown in **Supplementary Table 2**, the number of *T. pyogenes* was significantly higher in the endometritis group compared to the healthy group at 21 DPP in uterus (N.D. vs. 4.64 ± 0.23 log cfu/ml, $p < 0.05$) and vagina (2.62 ± 0.76 vs. 5.03 ± 0.27 log cfu/ml, $p < 0.05$), and at 35 DPP in the vagina (2.62 vs. 4.37 ± 0.73 log cfu/ml $p < 0.05$).

Differential Abundance Analysis

To verify the difference in genera between the healthy and endometritis groups, differential abundance analysis was performed using ALDEx2, which estimates the composition of biological features in the sample based on the number of reads.

Except for duplications, in total 10 of the bacterial genera were found to be different between the healthy and endometritis groups in uterine samples (**Figure 4**). The healthy group had higher abundances of genus *Enterococcus* at 0 DPP, and genus *Clostridium* at 35 DPP ($p < 0.024$ and $p < 0.045$, respectively). Genera differences were most detected at 28 DPP: *Prevotella*, *Bifidobacterium*, *Vibrio*, *Streptococcus*, *Pseudoalteromonas*, *Peptoniphilus*, and *Enterococcus* were significantly increased in the healthy group with $p < 0.0005$ to $p < 0.049$, whereas *Phascolarctobacterium* and *Eubacterium* showed a significant increase in the endometritis group ($p < 0.044$ and $p < 0.045$, respectively).

For vaginal samples, genus *Histophilus* and an unclassified genus of the family *Mogibacteriaceae* were found to be significantly increased in the endometritis group at pre-calving ($p < 0.048$ and $p < 0.047$, respectively), whereas unclassified genera of the families of *Lachnospiraceae* and *Ruminococcaseae* were significantly increased in the healthy group at 35 DPP ($p < 0.029$ and $p < 0.043$, respectively). There were no significant differences in the genera between healthy and endometritis groups at 0–28 DPP for vaginal samples, and 7–21 DPP for uterine samples.

Correlation of Uterine Bacterial Communities

The correlations between the top 20 abundant bacterial genera of the uterine and vaginal microbiota were analyzed mixing both healthy and endometritis groups (**Figure 5A**). There was a significantly positive correlation between *Trueperella* and *Helcococcus* ($r = 0.875$; $p < 0.001$) but significantly negative correlations with *Corynebacterium*, *Porphyromonas*, *Lactobacillus*, *Streptococcus*, unclassified genus of the family *Lachnospiraceae*, and unclassified genus of the order *Clostridiales*. The genus *Bifidobacterium* was shown to have strong positive correlations with the genera *Prevotella* and *Enterococcus* ($r = 0.964$ and 0.954, respectively; $p < 0.001$), and moderately positive correlations with *Lactobacillus* and *Streptococcus* ($r = 0.519$ and 0.513, respectively; $p < 0.01$). In the vaginal microbiota, there were statistically significant positive correlations between the genera *Trueperella* and *Helcococcus* ($r = 0.816$; $p < 0.001$),

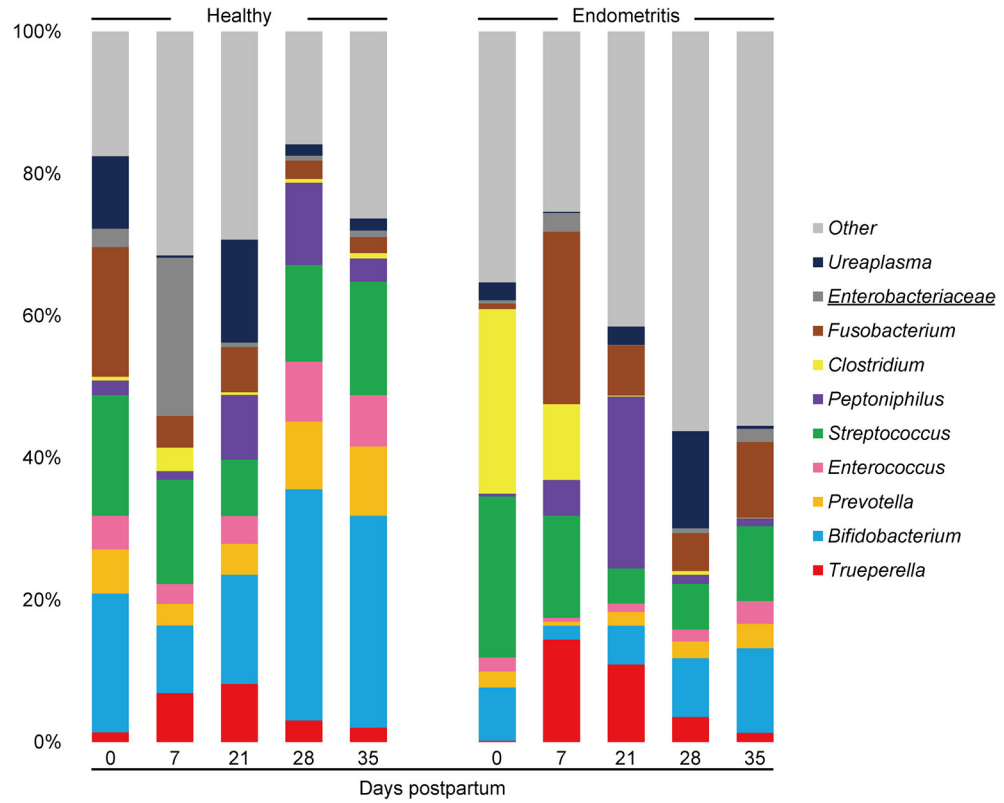


FIGURE 3 | Stacked bar chart showing mean relative abundance of 10 most abundant genera in uterine microbiota of healthy and endometritis cows. The underlined legends indicate classification only at the family level with genus not precisely defined.

as well as between *Trueperella* and *Peptoniphilus* ($r = 0.679$; $p < 0.001$), whereas *Trueperella* showed moderately negative correlations with unclassified genus of the family *Lachnospiraceae* and *Ruminococcaceae* ($r = -0.202$ and -0.275, respectively; $p < 0.01$) (**Figure 5B**). The genus *Ureaplasma* was the most predominant and negatively correlated with many other bacteria.

LEfSe Analysis on the Effect of Birth

For healthy cows, LEfSe analysis resulted in five genera which were significantly discriminative between "before birth" group (Pre) and "after birth" group (0 DPP) (>3 log 10 LDA score, $p < 0.05$) (**Supplementary Figure 4**). Genera *Helcococcus*, *Blautia*, and *Klebsiella* were represented in "before birth" group, while genus *Sutterella* and an unclassified genus of family *Alcaligenaceae*, were represented in "after birth" groups. By contrast, nine identified bacterial taxa including genera *Corynebacterium*, *Helcococcus*, an unclassified genus of the family *Veillonellaceae* were unique to "before birth" group of endometritic cows (>4 log 10 LDA score, $p < 0.05$).

DISCUSSION

Postpartum dairy cows are at high risk for a variety of diseases. There is some evidence that many risk factors are closely intertwined with endometritis, such as a negative energy balance, physiologic uterine inflammation, and the presence of bacteria in the environment and inside the genital tract (1, 31). In this study, we compared both uterine and vaginal microbiota of primiparous cows during early DPP using 16S rRNA gene-based metagenomic analysis in groups of cows with and without endometritis.

This study showed that the uterine microbiota at 7 DPP contained the lowest bacterial richness regardless of the presence or absence of endometritis. Uterine mucosa became rapidly colonized with environmental bacteria after delivery (8), and the number of PMNs also increased with the initial host immune response, peaking at 14 DPP (32). Therefore, these low microbial diversities at 7 DPP may be as a result of physiological inflammation which expel detritus. In order to understand the pathogenesis of endometritis it is important to clarify which pathogenic bacteria can consequently colonize the endometrium after 7 DPP. In contrast to the uterine microbiota, the vaginal microbiota had the highest richness before calving and decreased with time. Unlike humans, whose vaginal flora is usually dominated by genus *Lactobacillus*, those of healthy cows is composed of a variety of bacteria depending on the estrous cycle (33, 34). Therefore, it was expected that the bacterial composition of the vagina would not yet be fully restored to its prepartum state at 35 DPP. The diversity of the vaginal microbiota was higher in the endometritis group than in the

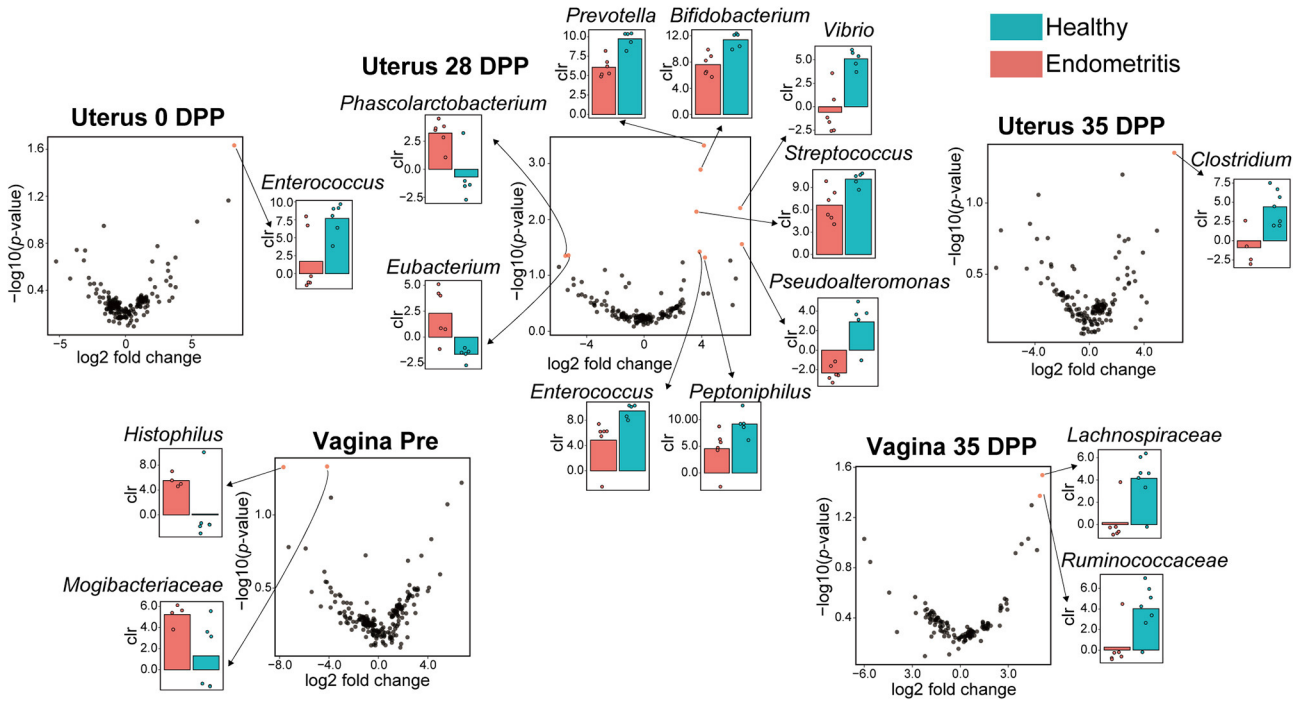


FIGURE 4 | Volcano plot showing significantly different genera between the healthy and endometritis groups at each of the indicated days postpartum. Red plots represent significant hits ($p < 0.05$). The centered log ratio (clr) values of each bacterial genus for healthy (blue) and endometritis (red) groups, which correspond to the red plots, are shown on the bar chart.

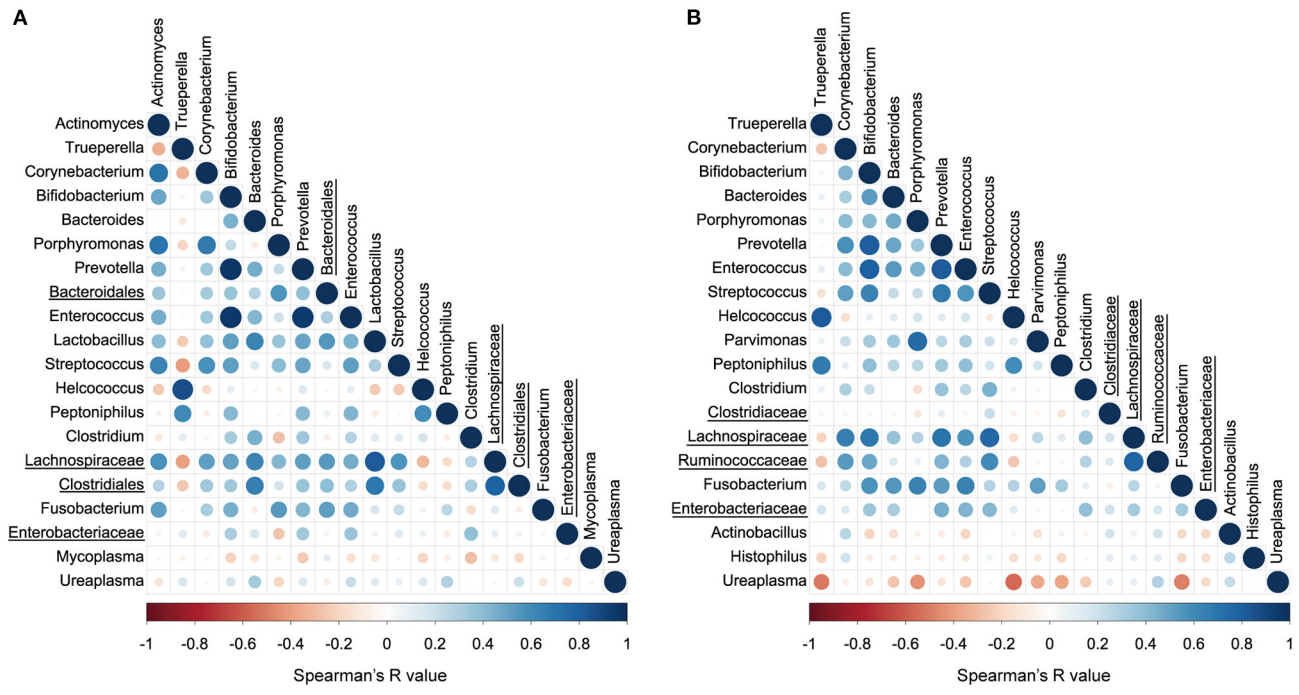


FIGURE 5 | Correlation plot of top 20 abundant bacterial genera in uterine **(A)** and vaginal microbiota **(B)**. The colors and circle sizes depict the degree of Spearman's rank correlation coefficient (rho) and *p*-value, respectively. The underlined legends indicate classified only at the order level or family level with the genus not precisely defined.

healthy group at 35 DPP, however there was no difference in beta-diversities (PERMANOVA $p = 0.22$).

The uterine and vaginal microbiota did not share more than 20.1% of all ASVs within the coverage period (from 0 to 35 DPP). Previous reports have shown that the postpartum uterine and vaginal microbiota shared 9.2% core OTUs (20), suggesting that each of the two compartments may be composed of independent populations of bacteria. The number of ASVs in the uterine and vaginal microbiota of healthy cows on 0 DPP was 207 and 287, respectively, whereas those of endometritis group were 414 and 173, suggesting that the presence of a large number of bacteria in the uterus at 0 DPP may have an effect on subsequent prognosis.

At the genus level, *Bifidobacterium* and *Streptococcus* were predominant in the uterine microbiota and *Ureaplasma* was predominant in the vaginal microbiota of healthy primiparous cows in this study. These results are consistent with the previous study which sampled both primiparous and multiparous cows using 16S rRNA sequence analysis (35). In contrast, cows diagnosed as suffering from endometritis at 35 DPP possessed *Clostridium* dominated uterine microbiota at 0 DPP, and *Trueperella* dominated uterine microbiota on 7 and 21 DPP, and the same trend was observed in the results estimated by the culture method in this study. Pascottini et al. reported that the relative abundance of *Trueperella* in the uterine microbiota estimated by 16S rRNA sequence analysis was increased in cows with clinical endometritis (18). In addition, *Trueperella* was isolated most frequently from the uterus in cows with clinical endometritis (detection rate: 43.5%), with a particularly high number of infected cows between 9 and 15 DPP (36). These results suggest that a high prevalence of *Trueperella* in the uterus is the most important risk factor for endometritis in primiparous cows.

Trueperella was significantly positively correlated with *Helcococcus*, but was negatively correlated with six genera including *Lactobacillus*. These results indicate that pathogenic bacteria may interact with each other to cause inflammation, and that building an environment to encourage bacteria that are generally recognized as beneficial, such as *Lactobacillus*, may be one possible solution to reduce the abundance of *Trueperella* and *Helcococcus*. Probiotic therapies have been recognized to exhibit a beneficial effect on bacterial vaginosis in human medicine (37), indicating that administering potentially beneficial bacteria to the reproductive tract of cows may help preventing reproductive diseases.

Unique biological marker candidates within the uterine and vaginal microbiota characterizing the healthy and endometritis groups were identified in this study. In the vaginal microbiota, *Histophilus* and *Mogibacteriaceae* were unique to the pre-calving period of the endometritis group. Deng et al. reported that these two bacteria were frequently abundant in vaginal samples of non-pregnant cows by using the random forest predictive model (38). These results suggest that the abundance of these two bacteria in the vagina of pre-calving cows may be associated with the development of endometritis. In this study, *Ruminococcaceae* and *Lachnospiraceae* were more common in the vagina of healthy cows at 35 DPP. In a previous report, the vaginal microbiota of healthy cows at 7 DPP was dominated by *Ruminococcaceae*, and *Lachnospiraceae* was also a common bacterium in healthy cows (17). Other studies have shown that *Ruminococcaceae* and *Lachnospiraceae* were present in the rumen of pre-weaned calves at 2.45 and 1.60%, respectively, and in the feces at 3.04 and 2.90%, respectively, and that their occupancy increases with growth (39, 40). Since feces to vagina microbiota transfer has been shown by previous studies (16), oral administration of these bacteria (*Ruminococcaceae* and *Lachnospiraceae*) to heifers prior to fertilization may transfer to the vagina via feces and contribute to the establishment of pregnancy, calving, and subsequent stable uterine recovery.

It was found that *Enterococcus* at 0 DPP, 7 bacterial genera including *Bifidobacterium* at 28 DPP, and *Clostridium* at 35 DPP are characteristic bacteria in the uterus of the healthy group. On the other hand, *Phascolarctobacterium* and *Eubacterium* were characteristic bacteria for uterine samples of the endometritis group. *Phascolarctobacterium* induces inflammation in humans, and its high prevalence has been reported to be associated with colorectal cancer (CRC) (41). *Eubacterium* has been generally recognized as a marker of anti-inflammation in humans. However, a recent study indicated that this bacterium may also be implicated in CRC development (42).

The populations of the vaginal microbiota of cows that had given birth was clearly different from those of before birth. However, it is not clear how this altered vaginal microbiota will change following structural and functional recover of both genital tract and ovaries. Previous studies have reported the inclusion of dystocia as a risk factor for metritis and endometritis (43, 44). In the present study, the presence of endometritis tended to be associated with the dystocia score (Mann-Whitney *U*-test, $p = 0.057$), suggesting that dystocia was indirectly associated with bacterial composition observed in endometritis group in this study.

In the current study, we analyzed microbiota of both uterus and vagina in the early postpartum period and identified unique bacterial genera that are characteristic of healthy and endometritis groups. Since the abundance and correlation between pathogenic bacteria, including *Trueperella*, exhibited similar trends to previous reports which did not discriminate primiparous and multiparous cows, the effect of parity on pathogenic bacterial colonization dynamics is thought to be small. However, in this study, differentiating primiparous cows has allowed us to characterize the bacterial dynamics in healthy cows which has not been previously reported. These findings could be useful for predicting endometritis and for developing prevention or treatment strategies. However, there are several limitations in our study: all primiparous cows came from a single farm, the sample size was small, and clinical and sub-clinical endometritis could not be analyzed separately. Therefore, our findings may be difficult to extrapolate to different regions or different cattle types. All primiparous cows used in this study was received AI or ET to be pregnant (no natural breeding), therefore the effect of the penile microbiota of the bull, which may be transferred by natural breeding, was not taken into account. The presence of microbiota in semen has become clearer in recent years (45). In this study, no single type semen was used in AI, but the microbiota in each semen was not analyzed. The complex

interactions (e.g., natural breeding, AI or ET, the microbiota of the semen used in AI, the degree of recovery of the uterus and ovaries after parturition) should be considered when assessing reproductive tract microbiota.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://www.ddbj.nig.ac.jp/ BioProject: PRJDB11963.

ETHICS STATEMENT

The animal study was reviewed and approved by Rakuno Gakuen University under protocol number VH16C7.

AUTHOR CONTRIBUTIONS

HK, TS, and MU: conceptualization. HK: methodology, investigation all experiments, and writing original draft. HK and TS: sample collection. HK: bacterial culture. HK and SH: 16S rRNA analysis. KO, MT, and MU: funding acquisition and resources. SK and YT: supervision. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.736996/full#supplementary-material

REFERENCES

1. Sheldon IM, Cronin J, Goetze L, Donofrio G, Schuberth HJ. Defining postpartum uterine disease and the mechanisms of infection and immunity in the female reproductive tract in cattle. *Biol Reprod.* (2009) 81:1025–32. doi: 10.1095/biolreprod.109.077370
2. Gilbert RO, Shin ST, Guard CL, Erb HN, Frajblat M. Prevalence of endometritis and its effects on reproductive performance of dairy cows. *Theriogenology.* (2005) 64:1879–88. doi: 10.1016/j.theriogenology.2005.04.022
3. Kasimanickam R, Duffield TF, Foster RA, Gartley CJ, Leslie KE, Walton JS, et al. Endometrial cytology and ultrasonography for the detection of subclinical endometritis in postpartum dairy cows. *Theriogenology.* (2004) 62:9–23. doi: 10.1016/j.theriogenology.2003.03.001
4. Williams EJ, Fischer DP, Pfeiffer DU, England GC, Noakes DE, Dobson H, et al. Clinical evaluation of postpartum vaginal mucus reflects uterine bacterial infection and the immune response in cattle. *Theriogenology.* (2005) 63:102–17. doi: 10.1016/j.theriogenology.2004.03.017
5. Zinsstag J, Schelling E, Waltner-Toews D, Tanner M. From "one medicine" to "one health" and systemic approaches to health and well-being. *Prev Vet Med.* (2011) 101:148–56. doi: 10.1016/j.prevetmed.2010.07.003
6. World Health Organization. *Antimicrobial Resistance: Global Report on Surveillance*. Geneva: World Health; Organization (2014).
7. Alipour MJ, Jalanka J, Pessa-Morikawa T, Kokkonen T, Satokari R, Hynonen U, et al. The composition of the perinatal intestinal microbiota in cattle. *Sci Rep.* (2018) 8:10437. doi: 10.1038/s41598-018-31494-3
8. Foldi J, Kulcsar M, Pecsi A, Huyghe B, de Sa C, Lohuis JA, et al. Bacterial complications of postpartum uterine involution in cattle. *Anim Reprod Sci.* (2006) 96:265–81. doi: 10.1016/j.anireprosci.2006.08.006
9. Sheldon IM, Dobson H. Postpartum uterine health in cattle. *Anim Reprod Sci.* (2004) 82–3:295–306. doi: 10.1016/j.anireprosci.2004.04.006
10. Dohmen MJW, Lohuis JACM, Huszenicza G, Nagy P, Gacs M. The relationship between bacteriological and clinical findings in cows with subacute/chronic endometritis. *Theriogenology.* (1995) 43:1379–88. doi: 10.1016/0093-691X(95)00123-P
11. Westermann S, Drillich M, Kaufmann TB, Madoz LV, Heuwieser W, A. clinical approach to determine false positive findings of clinical endometritis by vaginoscopy by the use of uterine bacteriology and cytology in dairy cows. *Theriogenology.* (2010) 74:1248–55. doi: 10.1016/j.theriogenology.2010.05.028
12. Hussain AM, Daniel RCW, O'Boyle D. Postpartum uterine flora following normal and abnormal puerperium in cows. *Theriogenology.* (1990) 34:291–302. doi: 10.1016/0093-691X(90)90522-U
13. Sheldon IM, Noakes DE, Rycroft AN, Dobson H. Effect of postpartum manual examination of the vagina on uterine bacterial contamination in cows. *Vet Rec.* (2002) 151:531–4. doi: 10.1136/vr.151.18.531
14. Wagener K, Prunner I, Pothmann H, Drillich M, Ehling-Schulz M. Diversity and health status specific fluctuations of intrauterine microbial communities in postpartum dairy cows. *Vet Microbiol.* (2015) 175:286–93. doi: 10.1016/j.vetmic.2014.11.017
15. Santos TM, Bicalho RC. Diversity and succession of bacterial communities in the uterine fluid of postpartum metritic, endometritic and healthy dairy cows. *PLoS One.* (2012) 7:e53048. doi: 10.1371/journal.pone.0053048
16. Bicalho MLS, Santin T, Rodrigues MX, Marques CE, Lima SF, Bicalho RC. Dynamics of the microbiota found in the vaginas of dairy cows during the transition period: associations with uterine diseases and reproductive outcome. *J Dairy Sci.* (2017) 100:3043–58. doi: 10.3168/jds.2016-11623
17. Miranda-CasoLuengo R, Lu J, Williams EJ, Miranda-CasoLuengo AA, Carrington SD, Evans ACO, et al. Delayed differentiation of vaginal and uterine microbiomes in dairy cows developing postpartum endometritis. *PLoS ONE.* (2019) 14:e0200974. doi: 10.1371/journal.pone.0200974
18. Pascottini OB, Van Schyndel SJ, Spricigo JFW, Rousseau J, Weese JS, LeBlanc SJ. Dynamics of uterine microbiota in postpartum dairy cows with clinical or subclinical endometritis. *Sci Rep.* (2020) 10:12353. doi: 10.1038/s41598-020-69317-z
19. Appiah M, Wang J, Lu W. Microflora in the reproductive tract of cattle: a review. *Agriculture.* (2020) 10:474. doi: 10.3390/agriculture10060232
20. Clemmons BA, Reese ST, Dantas FG, Franco GA, Smith TPL, Adeyosoye OI, et al. Vaginal and uterine bacterial communities in postpartum lactating cows. *Front Microbiol.* (2017) 8:1047. doi: 10.3389/fmicb.2017.01047
21. Nishijima S, Suda W, Oshima K, Kim SW, Hirose Y, Morita H, et al. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res.* (2016) 23:125–33. doi: 10.1093/dnares/dsw002
22. Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* (2011) 5:1571–9. doi: 10.1038/ismej.2011.41
23. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* (2019) 37:852–7. doi: 10.1038/s41587-019-0209-9
24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* (2016) 13:581–3. doi: 10.1038/nmeth.3869
25. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform.* (2014) 8:14. doi: 10.3389/fninf.2014.00014
26. Fernandes AD, Reid JNS, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* (2014) 2:15. doi: 10.1186/2049-2618-2-15
27. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes.* (2020) 6:60. doi: 10.1038/s41522-020-00160-w

28. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* (2011) 12:R60. doi: 10.1186/gb-2011-12-6-r60
29. Mitsuoka T. Cultivation of intestinal bacteria. In: Mitsuoka T, editor. *Intestinal Bacteria and Health*. Tokyo: Harcourt Brace Jovanovich Japan. (1978). p. 19–36.
30. Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. *Nucleic Acid Techniques in Bacterial Systematics*. Chichester: John Wiley and Sons (1991). p. 115–75.
31. Bradford BJ, Yuan K, Farney JK, Mamedova LK, Carpenter AJ. Invited review: Inflammation during the transition to lactation: new adventures with an old flame. *J Dairy Sci.* (2015) 98:6631–50. doi: 10.3168/jds.2015-9683
32. Pascottini OB, Van Schyndel SJ, Spricigo JFW, Carvalho MR, Mion B, Ribeiro ES, et al. Effect of anti-inflammatory treatment on systemic inflammation, immune function, and endometrial health in postpartum dairy cows. *Sci Rep.* (2020) 10:5236. doi: 10.1038/s41598-020-62103-x
33. Matsumoto A, Yamagishi Y, Miyamoto K, Oka K, Takahashi M, Mikamo H. Characterization of the vaginal microbiota of Japanese women. *Anaerobe.* (2018) 54:172–7. doi: 10.1016/j.anaerobe.2018.10.001
34. Quereda JJ, Barba M, Moce ML, Gomis J, Jimenez-Trigos E, Garcia-Munoz A, et al. Vaginal microbiota changes during estrous cycle in dairy heifers. *Front Vet Sci.* (2020) 7:371. doi: 10.3389/fvets.2020.00371
35. Jeon SJ, Vieira-Neto A, Gobikrushanth M, Daetz R, Mingoti RD, Parize AC, et al. Uterine microbiota progression from calving until establishment of metritis in dairy cows. *Appl Environ Microbiol.* (2015) 81:6324–32. doi: 10.1128/AEM.01753-15
36. Wagener K, Grunert T, Prunner I, Ehling-Schulz M, Drillich M. Dynamics of uterine infections with *Escherichia coli*, *Streptococcus uberis* and *Trueperella pyogenes* in post-partum dairy cows and their association with clinical endometritis. *Vet J.* (2014) 202:527–32. doi: 10.1016/j.tvjl.2014.08.023
37. Wang Z, He Y, Zheng Y. Probiotics for the treatment of bacterial vaginosis: a meta-analysis. *Int J Environ Res Public Health.* (2019) 16, 3859. doi: 10.3390/ijerph16203859
38. Deng F, McClure M, Rorie R, Wang X, Chai J, Wei X, et al. The vaginal and fecal microbiomes are related to pregnancy status in beef heifers. *J Anim Sci Biotechnol.* (2019) 10:92. doi: 10.1186/s40104-019-0401-2
39. Jami E, Israel A, Kotser A, Mizrahi I. Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* (2013) 7:1069–79. doi: 10.1038/ismej.2013.2
40. Meale SJ Li S, Azevedo P, Derakhshani H, Plaizier JC, Khafipour E, et al. Development of ruminal and fecal microbiomes are affected by weaning but not weaning strategy in dairy calves. *Front Microbiol.* (2016) 7:582. doi: 10.3389/fmicb.2016.00582
41. Candela M, Turroni S, Biagi E, Carbonero F, Rampelli S, Fiorentini C, et al. Inflammation and colorectal cancer, when microbiota-host mutualism breaks. *World J Gastroenterol.* (2014) 20:908–22. doi: 10.3748/wjg.v20.i4.908
42. Wang Y, Wan X, Wu X, Zhang C, Liu J, Hou S. Eubacterium rectale contributes to colorectal cancer initiation via promoting colitis. *Gut Pathog.* (2021) 13:2. doi: 10.1186/s13099-020-00396-z
43. Gröhn YT, Erb HN, McCulloch CE, Saloniemi HS. Epidemiology of metabolic disorders in dairy cattle: association among host characteristics, disease, and production. *J Dairy Sci.* (1989) 72:1876–85. doi: 10.3168/jds.S0022-0302(89)79306-1
44. Dubuc J, Duffield TF, Leslie KE, Walton JS, LeBlanc SJ. Risk factors for postpartum uterine diseases in dairy cows. *J Dairy Sci.* (2010) 93:5764–71. doi: 10.3168/jds.2010-3429
45. Altmae S, Franasiak JM, Mandar R. The seminal microbiome in health and disease. *Nat Rev Urol.* (2019) 16:703–21. doi: 10.1038/s41585-019-0250-y

Conflict of Interest: HK, SH, KO, MT, and SK are employed by Miyarisan Pharmaceutical Co., Ltd.

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

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fvets.2021.838703

Mechanisms of Oogenesis-Related Long Non-coding RNAs in Porcine Ovaries Treated With Recombinant Pig Follicle-Stimulating Hormone

Haiguang Mao[1†], Lu Chen[2†], Rupo Bao[2], Shiqiao Weng[2], Mengting Wang[1], Ningying Xu[3], Lili Qi[1] and Jinbo Wang[1*]*

[1] School of Biological and Chemical Engineering, NingboTech University, Ningbo, China, [2] Ningbo Sansheng Biological Technology Co., Ltd., Ningbo, China, [3] College of Animal Science, Zhejiang University, Hangzhou, China

Reproductive efficiency is of significant importance in pork production for it has a great impact on economic success. Ovulation rate is an early component of reproduction efficiency of pigs, and it contributes to the upper limit of litter size. In this study, we used the newly developed recombinant pig follicle stimulating hormone (rpFSH) instead of traditional PMSG to increase ovulation rate of pigs in order to achieve higher litter size, for it was better at stimulating ovulation, and showed more cheaper and greener. However, relatively little is known about the underlying genetic bases and molecular mechanisms. Consequently, an experiment was carried out in ovaries of replacement gilts to screen the key genes and lncRNAs that affect the fecundity of pigs by RNA-seq technology. Twenty gilts were divided into two groups, including 10 rpFSH treatment pigs and 10 control animals. After slaughtering and collecting the phenotypic data, ovaries of five pigs in each group were selected for RNA-seq. Total RNA was extracted to construct the library and then sequence on an Illumina Hiseq 4000 system. A comprehensive analysis of mRNAs and long non-coding RNAs (lncRNAs) from 10 samples was performed with bioinformatics. The phenotypic data showed that rpFSH treatment groups had the higher ($P < 0.01$) ovarian weight and more mature follicles. The RNA-seq results showed that a total of 43,499 mRNAs and 21,703 lncRNAs were identified, including 21,300 novel lncRNAs and 403 known lncRNAs, of which 585 mRNAs and 398 lncRNAs ($P < 0.05$) were significantly differentially expressed (DE) between the two groups of rpFSH treatment group and controlled group. GO and KEGG annotation analysis indicated that the target genes of DE lncRNAs and DE mRNAs were related to prolactin receptor activity, mitophagy by induced vacuole formation, and meiotic spindle. Moreover, we found that NR5A2 (nuclear receptor subfamily 5, group A, member 2), a target gene of lncRNA MSTRG.3902.1, was involved in regulating follicular development, ovulation, and estrogen production. Our study provided a catalog of lncRNAs and mRNAs associated with ovulation of rpFSH treatment, and they deserve further study to deepen the understanding of biological processes in the regulation of ovaries of rpFSH treatment pigs.

Keywords: lncRNA, ovulation, porcine ovary, rpFSH, reproduction

OPEN ACCESS

Edited by:
Hasan Riaz,
COMSATS Institute of Information Technology, Pakistan

Reviewed by:
Sayed Haidar Abbas Raza,
Northwest A&F University, China
Rajwali Khan,
University of Agriculture, Pakistan

***Correspondence:**
Lili Qi
qll@nbt.edu.cn
Jinbo Wang
wjb@nbt.edu.cn

†These authors share first authorship

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

Received: *18 December 2021*
Accepted: *31 December 2021*
Published: *24 February 2022*

Citation:
Mao H, Chen L, Bao R, Weng S, Wang M, Xu N, Qi L and Wang J (2022) Mechanisms of Oogenesis-Related Long Non-coding RNAs in Porcine Ovaries Treated With Recombinant Pig Follicle-Stimulating Hormone. Front. Vet. Sci. 8:838703. doi: 10.3389/fvets.2021.838703

INTRODUCTION

Fecundity is of primary interest in pig husbandry for it plays a vital role in the efficiency of production (1). Litter size, such as total number (of piglets) born (TNB), is one of the most important reproductive traits, which is difficult to be improved by traditional selection because of its relatively low heritability (2). In female animals, the mature oocyte quantity and quality are the two main factors affecting fertility (3). Ovulation is the first determinant factor for litter size, and many reports have shown that selection according to ovulation numbers could significantly increase litter size in sows (4, 5). Thus, in current pig production industry, reproductive hormones are widely used to achieve estrous synchronization and maximum reproductive genetic potential by improving the ovulation rate (6).

At present, batch management production of sow is mainly divided into four parts, including synchronization of sexual cycles, synchronization of follicular development, synchronization of ovulation, and synchronous mating (6). Pregnant mare serum gonadotropin (PMSG) is the most widely used in the synchronization of follicular development (7). In this study, we used the newly developed recombinant pig follicle stimulating hormone (rpFSH) instead of traditional PMSG to increase ovulation rate of pigs in order to achieve higher litter size, for it is better at stimulating ovulation, and shows more cheaper and greener. However, little is known about the underlying genetic bases and molecular mechanisms of the role in ovulation by rpFSH.

The ovary of sows, the most important reproductive organ, is responsible for synthesizing and secreting sex hormones, which are necessary for maintaining the hormone levels and the normal reproductive cycles (8). Follicular formation, ovulation, and luteal formation and regression all occur in the ovaries, and these processes take place repeatedly during mammalian reproduction and regulate reproduction (9). Previous reports have shown that long non-coding RNAs (lncRNAs) are involved in ovarian processes and regulate fertility (10–12).

Long non-coding RNAs (lncRNAs) are from regions of the transcriptome with lengths $>$ 200 nucleotides without the capacity of encoding evident proteins (13). Numerous evidences have indicated that the lncRNAs played important roles in the regulation of gene expression by directly recruiting epigenetic complexes or affecting the transcription process (10, 14, 15). To be more specific, lncRNAs could recruit transcription factors to DNA, segregating micro-RNAs (miRNAs) and destabilizing messenger (m) RNA (16). Therefore, the genetic mechanisms of cell differentiation, cell cycle regulation, epigenetics, and dosage compensation are all involved in the protein inhibition by binding of lncRNAs to miRNAs or to proteins or by miRNAs titration (17). LncRNAs have been reported as important regulatory factors in a variety of biological processes including reproduction, but the regulatory mechanism of lncRNAs in biological processes is largely unknown (18). The effects of lncRNAs on animal reproduction traits had been studied in recent years (19–21). The previous studies demonstrated that lncRNAs played an important role in the regulation of pigeon ovulation and sheep fertility (20, 22). Hu et al. (11) identified the ovarian lncRNAs associated with prolificacy of Large White sows during the follicular and luteal phases of the estrous cycle, and found that lncRNAs in ovaries significantly influenced fertility of pigs. Liu et al. (12) identified the lncRNA and mRNA expression profiles for pig ovaries on days 0, 2, and 4 of the follicular periods in Duroc pigs, and found that lncRNA ENSSSCT00000034907 might play an important role in follicular development. Although several researches have focused on the lncRNA expression profile of pig ovarian tissues, none of these studies have interpreted regulatory networks of lncRNAs for regulation of exogenous hormones on follicular formation in sow production.

In the present study, we are the first to perform transcriptome analysis of ovaries in gilts treated with recombinant pig follicle-stimulating hormone by RNA sequencing. The purpose of this study was to reveal the potential role of the lncRNAs in oogenesis treat by rpFSH and further provide a new insight in molecular mechanisms involved in the ovulation and fecundity in pigs. Our data provide a basis to understand the functional role of lncRNAs in improving reproductive rate performance in pigs.

MATERIALS AND METHODS

Animal and Ovary Collection

Twenty young gilts with the same genetic background were divided into two groups (reFSH treatment group and control group) with 10 gilts in each group. All the experiment gilts were born in the same day, raised in the same breeding environment, and fed by the diet according to the Nutrient Requirements of Swine Eleventh Revised Edition 2012. At 210 days of age, altrenogest was administered uniformly for 18 consecutive days, and each pig was fed 20 mg per day. After 18 days, all the pigs were stopped feeding altrenogest for 2 days. The guilts of rpFSH treatment group were injected 1,000 IU per pig, and the control group was injected the same volume of normal saline. The rpFHS (recombinant pig follicle-stimulating hormone) is a new reproductive hormone jointly developed by us and Ningbo Sansheng Biological Technology Co., Ltd. It is a protein-like hormone expressed in CHO. We used the newly developed rpFSH instead of traditional PMSG to increase ovulation rate of pigs in order to achieve higher litter size, for it was better at stimulating ovulation, and showed more cheaper and greener.

The ovary samples of the selected pigs were collected at 234 days of age after removing surface follicles and immediately frozen in liquid nitrogen to isolate RNA. Five samples from each group were randomly selected for sequencing. All animal procedures were approved by the Animal Welfare Committee of Zhejiang University.

RNA Isolation, Library Preparation, and Sequencing

The total RNA was isolated and purified from ovaries by TRIzol reagent (Invitrogen, Carlsbad, CA, USA) by the manufacturer's instructions. The amount and purity of the RNA samples were quantified by the NanoDrop ND-2000C (Thermo, USA). The extracted RNA integrity was measured by Agilent 2100 with RIN number $>$ 7.0. Approximately 5 μg of the extracted total RNA was used to remove rRNA by the instructions of the Ribo-Zero™

rRNA Removal Kit (Illumina, San Diego, USA). After removing the rRNAs, the rest of RNAs were fragmented into small pieces by divalent cations with high temperature. Afterwards, the cleaved RNA fragments were reversed into cDNA, and it was used to synthesize the U-labeled second-stranded DNAs with RNase H, *E. coli* DNA polymerase I, and dUTP. The average insert size of the final cDNA library was about 300 ± 50 bp. Finally, Illumina Hiseq 4000 (LC Bio, China) was used for the paired-end sequencing according to the recommended protocol of the apparatus. The RNA sequencing data has been uploaded in GEO with the accession number GSE192605.

Quality Control and Mapping

The low-quality reads, including low quality bases, adaptor contamination, and undetermined bases, were removed by Cutadapt. Sequence quality was then verified by FastQC. Hisat2 (23) and Bowtie2 (24) were used to map the reads to the genome of pigs (*Sus scrofa*) in NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/025/GCF_000003025.6_Sscrofa11.1/GCF_000003025.6_Sscrofa11.1_genomic.fna.gz). The mapped reads for each sample were then assembled by StringTie. All the transcripts from pig ovaries was combined to reconstruct a comprehensive transcriptome by a Perl script. Ballgown (25) and StringTie (26) were performed to estimate the expression levels of all the transcripts after the final transcriptome were generated.

Identification of lncRNAs

First of all, transcripts shorter than 200 bp or overlapped with known mRNAs were discarded. CNCI (27) and CPC (28) were applied to predict transcripts with coding potential. All the transcripts of CNCI scores < 0 and CPC score < -1 were removed. The rest transcripts were considered as lncRNAs.

Different Expressed mRNAs and lncRNAs Analysis

The mRNA and lncRNA expression levels were calculated with FPKM by StringTie. The DE mRNAs and DE lncRNAs were selected with the statistical significance ($P < 0.05$) and with log2 (fold change) < -1 or log2 (fold change) > 1 by R package-Ballgown.

Prediction of Target Gene and Functional Analysis of lncRNAs

The cis-target genes of lncRNAs were predicted to explore the function of lncRNAs, which were likely play a cis role on the neighboring target genes. In this study, the coding genes in 100,000 downstream and upstream were selected using python script (29). Afterwards, functional analysis of the target genes of lncRNAs was performed by the BLAST2GO (30).

GO and KEGG Enrichment Analysis

Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed to explore the biological processes, which might contribute to further understanding the biological functions of DE lncRNAs in pigs treated with rpFSH.

RNA-seq Result Validation by qRT-PCR

Six lncRNAs (MSTRG.16871.1, MSTRG.29090.1, MSTRG.34178.2, MSTRG.30735.1, MSTRG.33729.1, and MSTRG.33864.1) and six mRNAs (BET1L, RSAD2, CCR1, DRC7, LRRC46, and CFAP161) representing differential expression levels of RNA-seq from the 10 pig ovaries were randomly selected to perform qRT-PCR. The qRT-PCR was performed by the ABI Step One Plus system (Applied Biosystem, Carlsbad, CA, USA) with SYBR Premix Ex Taq kit (TaKaRa, Dalian, China), and the primers were shown in **Supplementary Table S1**. Relative gene expression levels were quantified and normalized by *β-actin* gene using $2^{-\Delta\Delta Ct}$ method with three independent biological replicates. All the measurements were performed in triplicate. Bonferroni correction method was used in the multiple comparison.

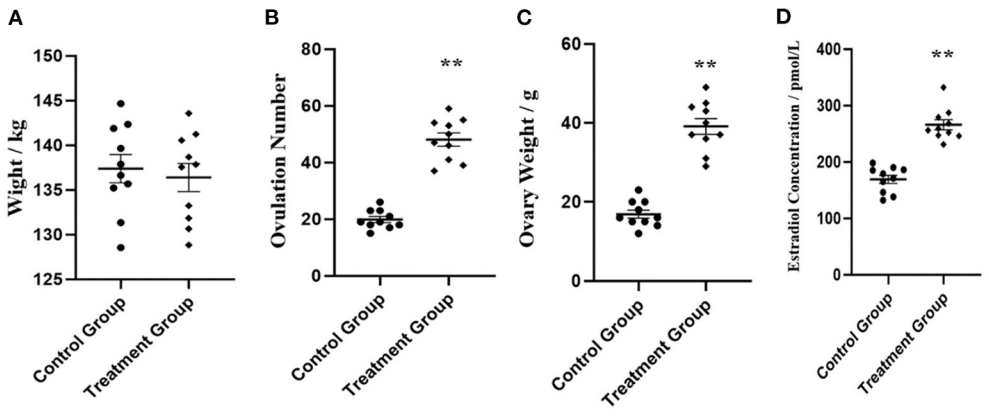


FIGURE 1 | Phenotypic data analysis of experiment gilts between control group and rpFSH treatment group ($n = 10$). **(A)** Live weight analysis. **(B)** Ovulation number analysis of bilateral ovaries. **(C)** Ovary weight analysis of bilateral ovaries. **(D)** Estradiol concentration analysis in serum. $^{**}P < 0.01$.

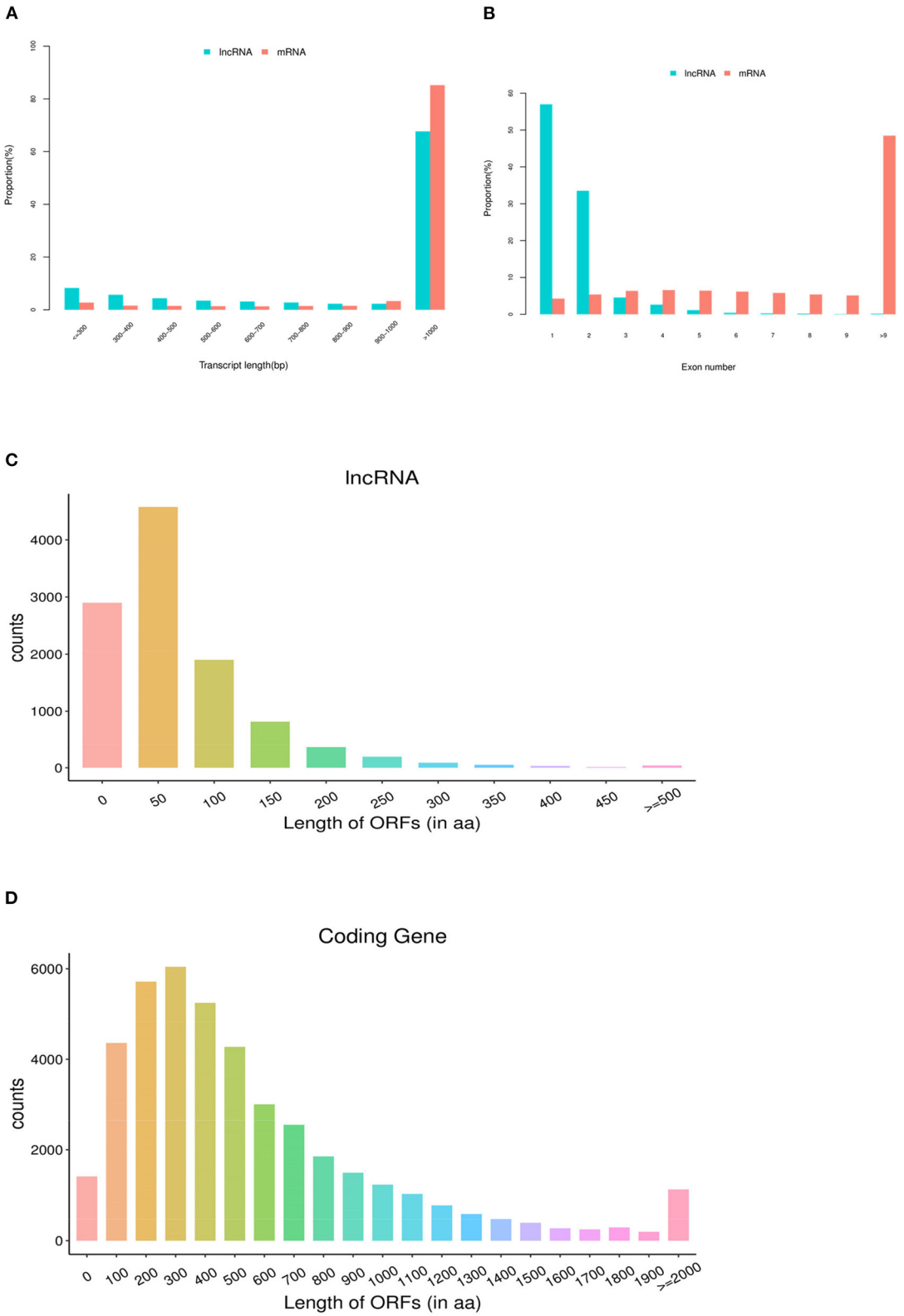


FIGURE 2 | Genomic features of lncRNAs in pig ovaries. **(A)** The transcript length distribution of lncRNAs and mRNAs. **(B)** The exon number distribution of lncRNAs and mRNAs. **(C)** The ORF length distribution of lncRNAs. **(D)** The ORF length distribution of mRNAs.

RESULTS

Phenotypic Data Analysis

Four phenotypic traits were collected, including live weight, ovulation number bilateral ovaries, ovary weight of bilateral ovaries, and estradiol concentration in serum ($n = 10$). As shown in **Figure 1**, no significant ($P > 0.05$) body weight was shown between the control group and rpFSH treatment group. As expected, the rpFSH treatment group showed significantly higher ($P < 0.01$) ovulation number of bilateral ovaries, ovary weight of bilateral ovaries, and estradiol concentration in serum.

Sequencing Date Summary

A total of 132.99 Gb raw date was obtained from ten libraries. In detail, 84853980, 87895596, 81995902, 93361996, and 96713942 raw reads were generated from the control group (Control1, Control2, Control3, Control4 and Control5); 88923972, 93496552, 83448796, 86664168, and 89271778 raw reads were

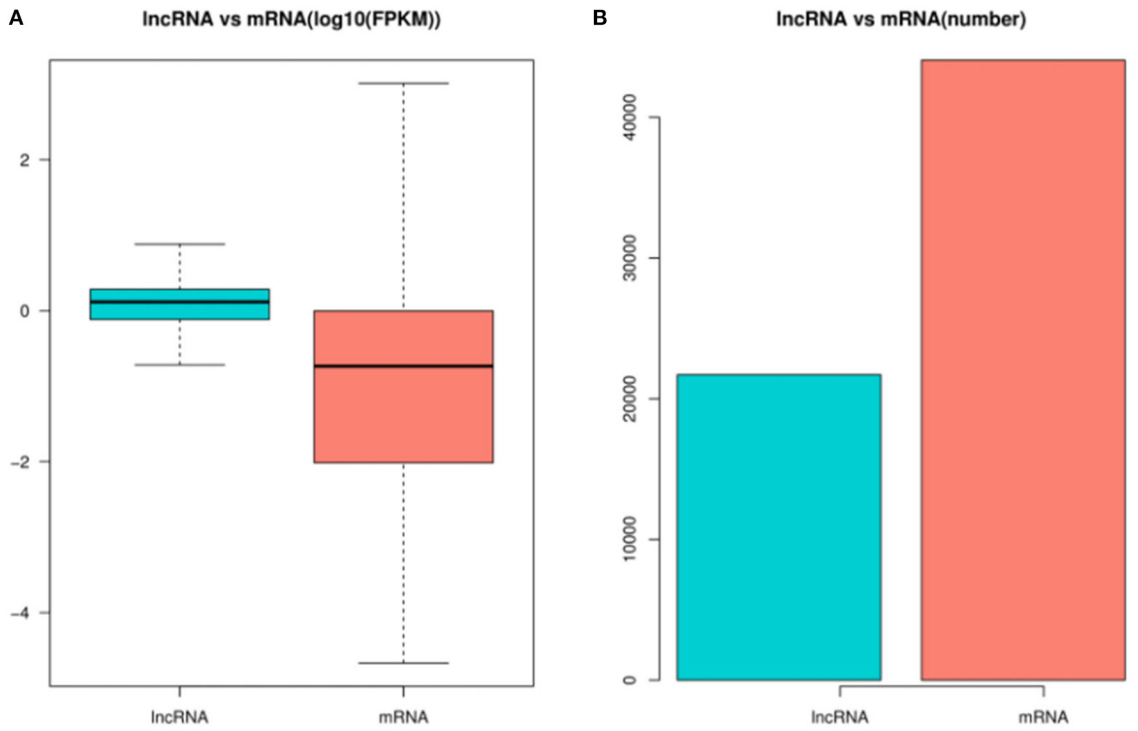


FIGURE 3 | The expression levels and amounts of lncRNAs and mRNAs. **(A)** Boxplots of lncRNA and mRNA expression levels (with log10 FPKM method) in the control group and treatment group. **(B)** The numbers of lncRNAs and mRNAs in pig ovaries in in the control group and treatment group.

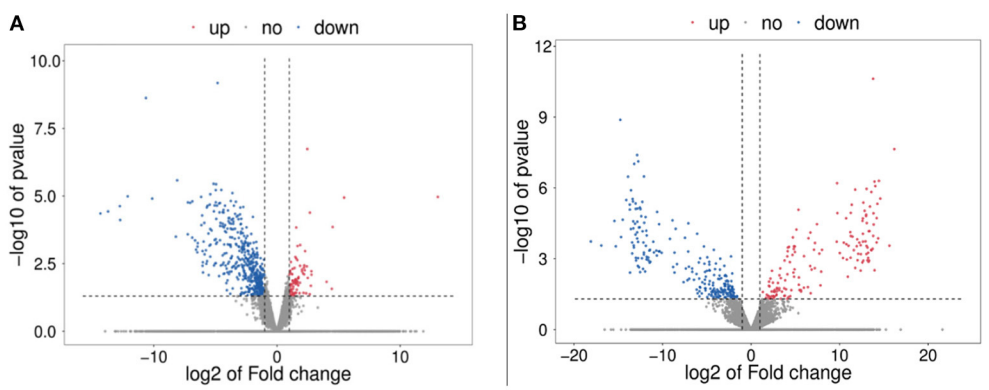


FIGURE 4 | Volcano plot of the differential expression of mRNAs and lncRNAs in pig ovaries between control group and treatment group. **(A)** Differential expression of mRNAs. The blue points denote significantly downregulated mRNAs, while the red points denote significantly upregulated mRNAs. **(B)** Differential expression of lncRNAs. The blue points denote significantly down-regulated lncRNAs, while the red points denote significantly upregulated lncRNAs.

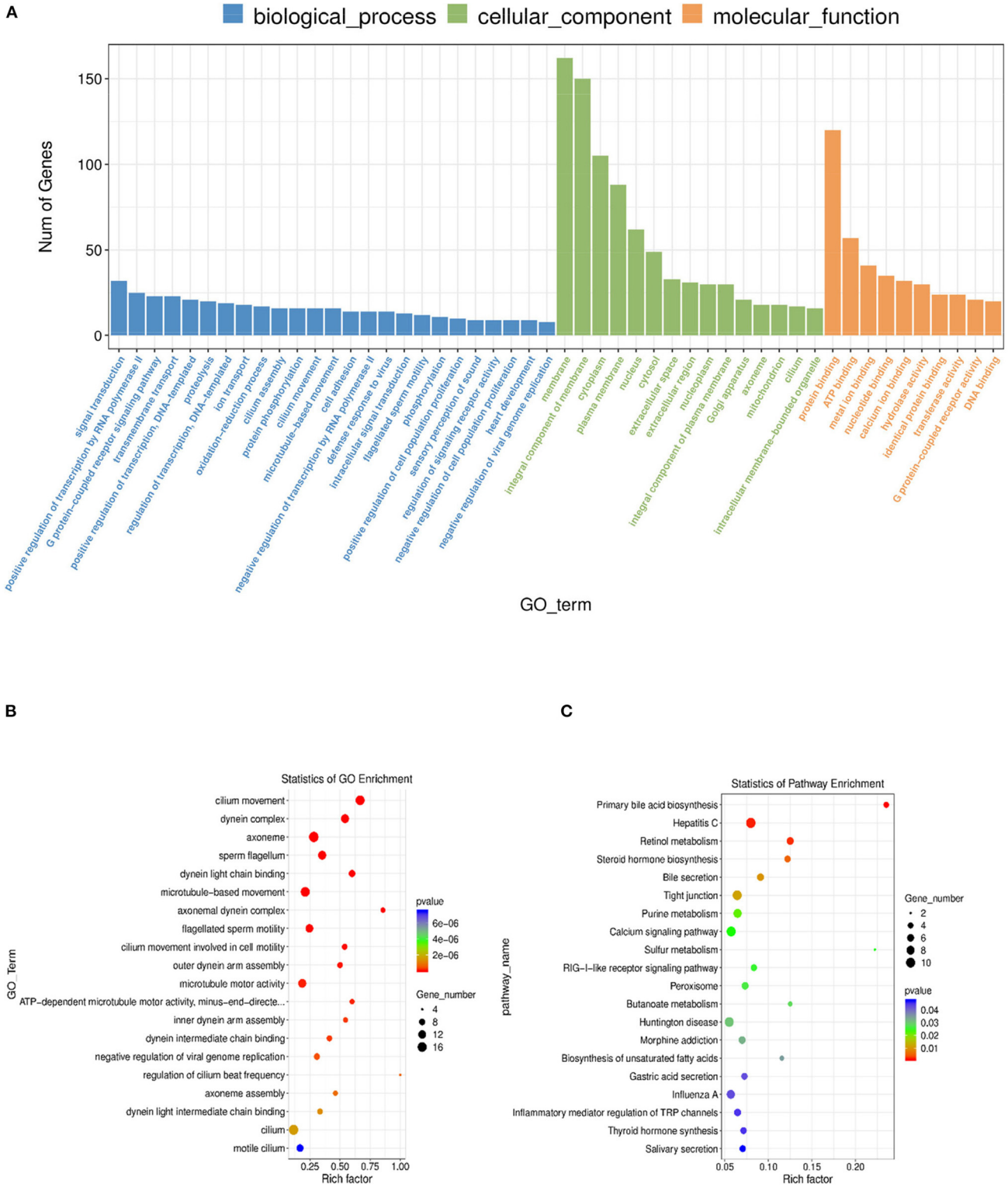


FIGURE 5 | GO and KEGG analysis of differential mRNA expression. **(A)** Histogram of GO enrichment of DE mRNAs. **(B)** Scatter plot of GO enrichment for DE mRNAs. **(C)** Scatter plot of KEGG enrichment for DE mRNAs.

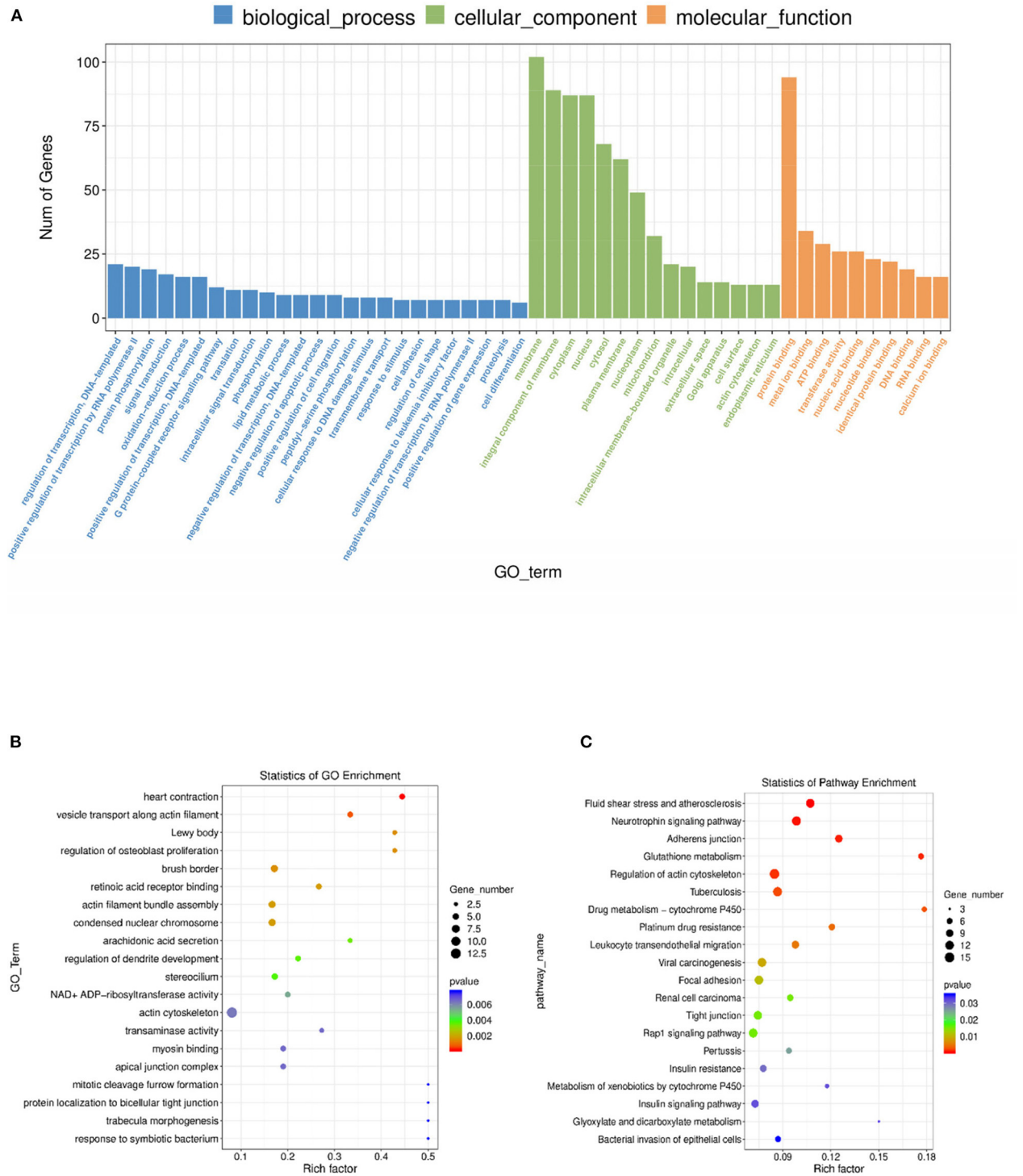


FIGURE 6 | GO and KEGG analysis of differentially lncRNA expression. **(A)** Histogram of GO enrichment of DE lncRNAs. **(B)** Scatter plot of GO enrichment for DE lncRNAs. **(C)** Scatter plot of KEGG enrichment for DE lncRNAs.

generated from the treatment group (Treat1, Treat2, Treat3, Treat4, and Treat5). All the raw reads were filtered to obtain the clean reads, which were mapped to the Cliv_1.0 version of the pig (*Sus scrofa*) genome sequence, with the mapping ratio ranging from 91.13 to 93.07%. The detailed dates are shown in **Supplementary Table S2**.

TABLE 1 | Differentially expressed lncRNA-gene pairs between high and low egg production groups.

Gene name	lncRNA transcript name	Cislocation(bp)	Pearson correlation coefficient
DOCK4	MSTRG.16790.2	100K	1
WDR49	MSTRG.8936.1	100K	1
ENSSSCG00000033862	MSTRG.23824.1	100K	1
SCP2	MSTRG.29550.1	100K	1
ENSSSCG00000033862	MSTRG.23822.2	100K	1
PXYLP1	MSTRG.8684.1	100K	−0.21
PXYLP1	MSTRG.8685.1	100K	−0.20
PHF21A	MSTRG.17683.2	100K	−0.20
PHF21A	MSTRG.17683.2	100K	−0.11

Identification of lncRNAs and mRNAs in Pig Ovaries

As shown in **Supplementary Table S3**, a total of 21,703 putative lncRNAs were identified from the 10 libraries, including 21,300 novel lncRNAs and 403 known lncRNAs. Regarding the genomic locations of the lncRNAs, 11,918 were intronic (54.91%), 605 were bidirectional (2.79%), 1,683 were sense (7.75%), 6,082 were intergenic (28.03%), and 1,415 were antisense lncRNAs (6.52%).

In this study, the average length of identified lncRNA transcripts is 2,411 bp, which shows shorter than 4,896 bp length of the mRNA transcripts (**Figure 2A**). Moreover, the number of exons of lncRNAs is 1.63 on average, which is less than that of mRNAs (11.81 on average).

As shown in **Figure 2B**, 95.09% of lncRNAs contain three or fewer exons, while 77.42% of mRNAs contain five or more exons. In addition, lncRNAs found in the present study showed shorter open reading frames (ORFs) than mRNAs of ovarian tissues in pigs (**Figures 2C,D**).

Identification of DE mRNAs and DE lncRNAs

In order to identify the mRNAs and DE lncRNAs between the control group and treatment group, we calculated the DE mRNAs and DE lncRNA expression levels with FPKM levels in pig ovaries. **Figure 3A** shows that the lncRNA expression levels were higher than mRNA expression levels in this study, while **Figure 3B** shows that the number of lncRNAs was less than that of mRNAs.

A total of 585 DE mRNAs (**Supplementary Table S4**) and 398 DE lncRNAs (**Supplementary Table S5**) were identified between the control group and treatment group. Compared with the control group, 85 mRNAs and 155 lncRNAs were significantly upregulated, while 500 mRNAs and 243 lncRNAs were downregulated. The volcano plot of the DE mRNAs and DE lncRNAs was shown in **Figure 4**.

Functional Enrichment of DE mRNAs

Gene Ontology (GO) was performed to analyze the main functions of the obtained DE mRNAs. A total of 2,417 GO terms with functional annotation information were enriched for 585 DE mRNAs. As shown in **Supplementary Table S6**, 354 GO terms significantly ($P < 0.05$) enriched in the GO analysis results of DE mRNAs. As shown in **Figures 5A–C**, the significantly enriched GO terms of DE mRNAs involved cilium movement, dynein complex, axoneme, sperm flagellum, and dynein light chain binding. KEGG pathway analysis showed 20 significantly ($P < 0.05$) enriched pathways, such as primary bile acid biosynthesis, hepatitis C, retinol metabolism, steroid hormone biosynthesis, and bile secretion. The detailed information was shown in **Supplementary Table S7**.

TABLE 2 | Co-enriched GO terms of DE lncRNA and DE mRNA.

GO term	GO function	*P*-value
Prolactin receptor activity	molecular_function	0.04
Brush border	cellular_component	0.00
Protein ADP-ribosylation	biological_process	0.02
Mitophagy by induced vacuole formation	biological_process	0.04
Meiotic spindle	cellular_component	0.02

Cis-Regulatory Roles of DE lncRNAs in Ovarian Tissues of Pigs

To further investigate the regulatory functions of the lncRNAs in the ovarian tissues of pigs, we forecasted the cis-regulated target genes of the differently expressed lncRNAs between the control group and treatment group. In this study, 62 potential lncRNA target genes were found, with 100 kbp as the cutoff (**Supplementary Table S8**). As shown in **Supplementary Table S9**, GO analysis revealed 220 significant ($P < 0.05$) GO terms based on the cis-regulated target genes. The differentially expressed lncRNA target genes were founded to be related with biological process including heart contraction, vesicle transport along actin filament, regulation of osteoblast proliferation, and actin filament bundle assembly. The main molecular function and cellular component categories were related to the lewy body, brush border, retinoic acid receptor binding, and condensed nuclear chromosome (**Figures 6A,B**). The KEGG analysis of DE lncRNAs revealed that the target genes of those lncRNAs were mainly enriched in fluid shear stress and atherosclerosis, neurotrophin signaling pathway, adherens junction, glutathione metabolism, and regulation of actin cytoskeleton (**Figure 6C**, **Supplementary Table S10**). Based on the prediction of DE lncRNA-gene pairs in cis-regulation, the first 5 and the last 5 lncRNA-gene pairs were listed in **Table 1** by the Pearson correlation coefficient, and the regulation directions of the first 5 lncRNA-gene pairs showed the same, while the last 4 pairs were opposite.

Co-enriched GO Terms of DE lncRNAs and mRNAs

In order to investigate the crucial pathways of rpFSH to gilt ovaries, a total of five significantly enriched GO terms

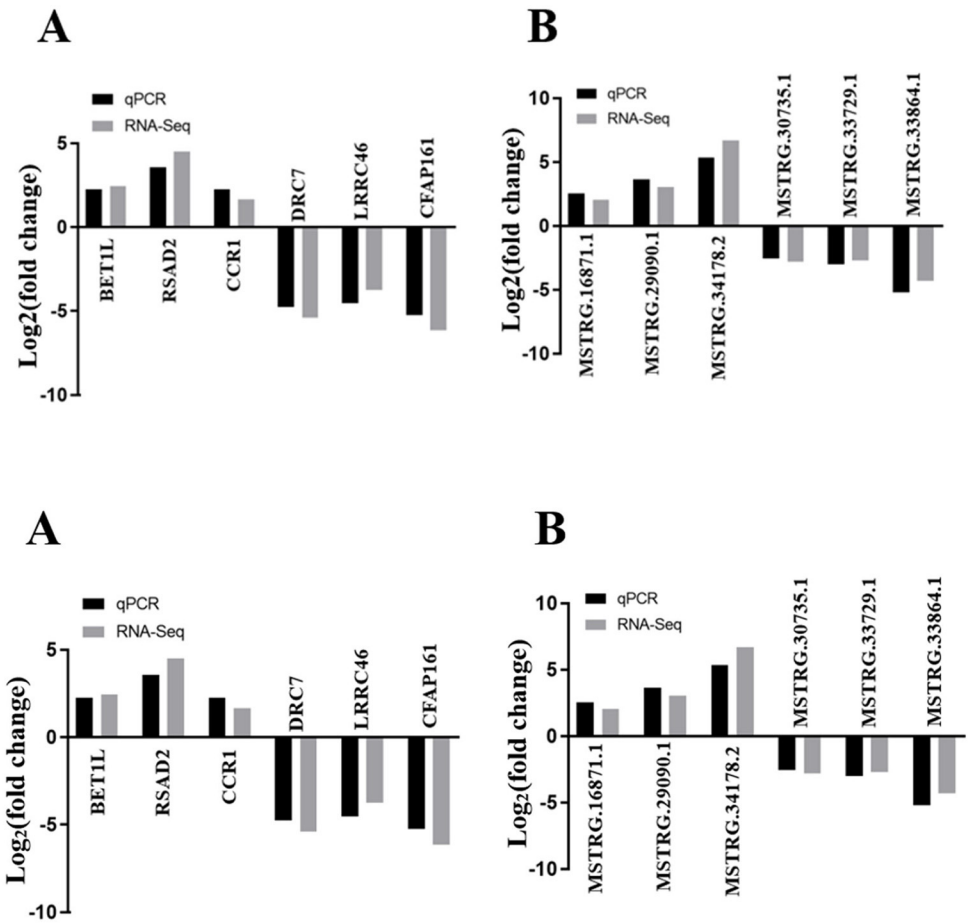


FIGURE 7 | The validation of RNA-seq by qRT-PCR ($n = 5$). **(A)** qRT-PCR validation of six mRNAs. **(B)** qRT-PCR validation of six lncRNAs.

were identified in both DE lncRNA target gene enrichment and DE mRNA enrichment (**Table 2**). The significantly co-enriched GO terms were involved in the prolactin receptor activity, brush border, protein ADP-ribosylation, mitophagy by induced vacuole formation, and meiotic spindle, of which one pathway was involved in molecular function, and two pathways were involved in cellular component and biological process, respectively.

DE lncRNAs and DE mRNA Validation by qRT-PCR

Six DE lncRNAs and six DE mRNAs were selected at random to validate the RNA-seq result by qRT-PCR. The relative fold changes of expression levels performed by qRT-PCR were consistent with the results of RNA-seq data (**Figure 7**), suggesting that the identification of transcripts and estimation of abundance were highly credible in this study. What is more, the relative mRNA expression level of NR5A2 in the rpFSH group was significantly higher ($P < 0.01$) than that in the control group (**Figure 8**).

DISCUSSION

The oogenesis and ovulation in mammals are relatively complex biological processes, which are well-coordinated and regulated by coding and non-coding RNAs (12). Ovulation number is a very crucial trait in pigs for it determines the maximum litter size (4, 5), which is one of the most important economic traits and is difficult to be improved by traditional selection because of its relatively low heritability (2). To achieve the aim of high ovulation rate in pigs, scientists had made great effort on the regulation mechanisms of animal reproduction, which had been greatly driven to perform related gene screening involved in the reproductive regulation of pigs and to mediate the process for increasing the litter size (1, 3, 9, 31). In the last few decades, the whole genomes of pigs had been continuously published, which contributed to facilitate the studies on the transcriptome in pigs (9, 10). In this study, gilts treated with rpFSH showed higher ovulation number and estrogen concentration, which might contribute to improving the litter size of pigs. Many researches have been reported that lncRNAs could regulated reproduction of pigs (10–12), but none of these studies have interpreted regulatory networks of lncRNAs for regulation of

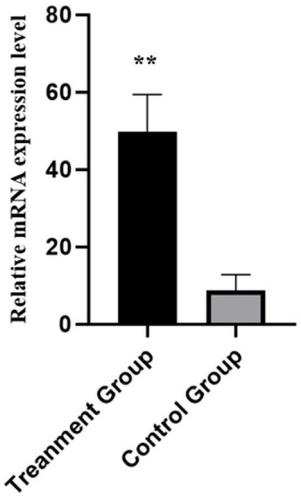


FIGURE 8 | Relative mRNA expression levels of pig *NR5A2* gene in ovaries between rpFSH treatment group and the control group ($n = 5$). **indicate significant differences ($P < 0.01$).

exogenous hormones on follicular formation in sow production. Therefore, in this study, we performed transcriptome analysis of ovaries in pigs treated with rpFSH by RNA sequencing. Finally, we identified 398 DE lncRNAs and 585 DE mRNAs in ovaries between the rpFSH treatment group and control group.

In the last decade, a number of researches had reported that lncRNAs played important roles in the oogenesis and ovulation in ovaries among different species, including pigs, cattle, mice, and sheep (11, 21, 32, 33). The present study is the first to report the transcriptome profiling of lncRNAs and mRNA in pig ovaries treated with rpFSH. Our sequencing results showed that the lncRNAs identified in the present study had shorter transcript lengths and fewer exons; this was consistent with the previous research results (9, 20), implying that the lncRNA sequencing result of this study was credible. The RNA-seq results showed that 32.32% of identified lncRNAs were shorter than 1,000 bp, while only 19.18% of identified mRNAs were shorter than 1,000 bp. In addition, the average expression levels of lncRNAs in this study were much higher ($P < 0.05$) than those of mRNAs in ovaries of pigs, suggesting that the lncRNAs in pig ovaries might play important roles in oogenesis and ovulation.

Many studies have found that numerous signaling pathways and regulatory mechanisms are taken part in the regulation of oogenesis and ovulation in animals (11, 21, 32, 33). In the current study, we performed GO terms and KEGG pathways analysis to further identify the biological functions of the target genes of DE mRNAs and DE lncRNAs related to oogenesis and ovulation in pig ovaries. The results revealed that both of these DE mRNAs and DE lncRNAs were participated in the regulation of protein binding, ATP binding, cell differentiation, and transcription by RNA polymerase II. It can be illustrated that ATP binding has been reported to participate in the oogenesis regulation (34). The RNA polymerase II was reported to be involved in the combinatorial control of Spo11 splicing, which is timely regulated during meiosis (35).

The previous studies have shown that the expression of lncRNAs can regulate the expression of the neighboring mRNAs by transcriptional repression or coactivation patterns and had high correlations with the expression levels with the adjacent genes (36, 37). In consequence, we speculated that there was a genetic mechanism that the lncRNAs could significantly affect the oogenesis and ovulation by mediating the putative regulation of the corresponding target mRNAs in pig ovaries. In the current study, the DE cis-target genes, which were located within 100 kb downstream and upstream of the 398 DE lncRNAs, were selected to predict the potential biological functions in the putative regulation of oogenesis and ovulation in pigs. The result suggested that the DE coding gene NR5A2 (nuclear receptor subfamily 5, group A, member 2) might be regulated by the DE lncRNA MSTRG.3902.1, and NR5A2 was significantly upregulated in rpFSH treatment group.

NR5A2, also known as liver receptor homolog-1 (LRH1), is an important orphan receptor, which belongs to the nuclear receptor subfamily NR5A (38). NR5A2 plays a significant role in somatic cell reprogramming, embryonic development, steroid hormone production, and follicle and oocyte development (39). In adult mammals, NR5A2 is mainly expressed in liver, intestine, and ovary tissues, especially in ovarian tissues where it is highly expressed (40, 41), indicating that NR5A2 might play an important role in the reproduction of female animals. NR5A2 gene knockout mice showed ovulation dysfunction and infertility, suggesting that NR5A2 is necessary for follicular development and ovulation in mammals (42, 43). The expression level of *NR5A2* gene in ovary was positively correlated with estrogen content (40), and *NR5A2* regulated porcine follicular estrogen secretion and granular cell apoptosis by targeting *CYP19A1* and *CYP11A1* genes (42). Moreover, the polymorphism of *NR5A2* gene showed significant association with litter size in Hu sheep (33). Our analysis result showed that the mRNA expression level of *NR5A2* in rpFSH treatment group was significantly higher ($P < 0.01$) than that of the control group; thus, we inferred that the NR5A2 might play an important role in the oogenesis and estrogen secretion. Therefore, we speculated that the *NR5A2* gene could be a candidate gene for further study in terms of how it affected oogenesis and ovulation in rpFSH treated pigs.

In conclusion, this study is the first comprehensive description of mRNA and lncRNA profiles of porcine ovaries treated with rpFSH. Several DE lncRNAs are revealed to be associated with ovulation number treated with rpFSH. Moreover, the DE lncRNAs identified in the present study could provide new insights for further understanding the mechanism of ovulation in pigs. The lncRNA MSTRG.3902.1 might play an important regulatory role in ovulation treated with rpFSH by affecting its potential target gene NR5A2. Therefore, lncRNA MSTRG.3902.1 might be a potential candidate lncRNA for regulating oogenesis in pigs treated with rpFSH, and more detailed studies should be carried out to verify the results.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192605.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Welfare Committee of Zhejiang University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

HM analyzed the data and drafted the manuscript. LC and RB collected the tissue samples and analyzed the data. MW performed the qRT-PCR. SW and NX provided suggestions for this study. LQ and JW conceived the project and designed the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

The current work was funded by Ningbo Science and Technology Innovation 2025 Project (NO. 2018B10095), Ningbo Major Science and Technology Project (NO. 2021Z112), and Talent Introduction Fund of NingboTech University (NO. 20211018Z0216).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.838703/full#supplementary-material

REFERENCES

1. Metodiev S, Thekkoot DM, Young JM, Onteru S, Rothschild MF, Dekkers JCM. A whole-genome association study for litter size and litter weight traits in pigs. *Livest Sci.* (2018) 211:87–97. doi: 10.1016/j.livsci.2018.03.004
2. Roehe R, Kennedy BW. Estimation of genetic parameters for litter size in Canadian Yorkshire and Landrace swine with each parity of farrowing treated as a different trait. *J Anim Sci.* (1995) 73:2959–70. doi: 10.2527/1995.73102959x
3. Sell-Kubiak E. Selection for litter size and litter birthweight in Large White pigs: maximum, mean and variability of reproduction traits. *Animal.* (2021) 15:100352. doi: 10.1016/j.animal.2021.100352
4. Ziadi C, Mocé LM, Laborda P,Blasco A,Santacreu AM. Genetic selection for ovulation rate and litter size in rabbits: estimation of genetic parameters and direct and correlated responses. *J Anim Sci*. (2013) 91:3113–20. doi: 10.2527/jas.2012-6043
5. Rosendo A, Druet T, Gogué J, Bidanel JP. Direct responses to six generations of selection for ovulation rate or prenatal survival in Large White Pigs. *J Anim Sci.* (2007) 85:356–64. doi: 10.2527/jas.2006-507
6. Breen SM, Rodriguez-Zas SL, Knox RV. Effect of altering dose of PG600 on reproductive performance responses in prepubertal gilts and weaned sows. *Anim Reprod Sci.* (2006) 95:316–23. doi: 10.1016/j.anireprosci.2005.12.007
7. Blitek A, Szymanska M, Pieczywek M, Morawska-Pucinska E. Luteal P4 synthesis in early pregnant gilts after induction of estrus with PMSG/hCG. *Anim Reprod Sci*. (2016) 166:28–35. doi: 10.1016/j.anireprosci.2016.01.001
8. Liu Y, Li M, Wang T, Guan J, Luo Z, Chen H, et al. Repertoire of porcine microRNAs in adult ovary and testis by deep sequencing. *Int J Biol Sci*. (2011) 7:1045–55. doi: 10.7150/ijbs.7.1045
9. Huang L, Yin ZJ, Feng YF, Zhang XD, Wu T, Ding YY. Identification and differential expression of microRNAs in the ovaries of pigs (Sus scrofa) with high and low litter sizes. *Anim Genet.* (2016) 47:543–51. doi: 10.1111/age.12452
10. Gao Y, Li SP, Lai ZY, Zhou ZH, Wu F, Huang YZ, et al. Analysis of long noncoding RNA and mRNA expression profiling in immature and mature bovine (Bos taurus) testes. *Front Genet*. (2019) 10:646. doi: 10.3389/fgene.2019.00646
11. Hu H, Jia Q, Xi J, Zhou B, Li Z. Integrated analysis of lncRNA, miRNA and mRNA reveals novel insights into the fertility regulation of large white sows. *BMC Genomics*. (2020) 21:636. doi: 10.1186/s12864-020-07055-2
12. Liu Y, Li M, Bo X, Li T, Ma L,Zhai T, et al. Systematic analysis of long non-coding RNAs and mRNAs in the ovaries of duroc pigs during different follicular stages using RNA sequencing. *Int J Mol Sci*. (2018) 19:1722. doi: 10.3390/ijms19061722
13. Aliaksandr AY, Igor VK. Long noncoding RNAs: a potential novel class of cancer biomarkers. *Front Genet*. (2015) 6:145–54. doi: 10.3389/fgene.2015.00145
14. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics.* (2013) 193:651–69. doi: 10.1534/genetics.112.146704
15. Mallory AC, Shkumatava A. LncRNAs in vertebrates: advances and challenges. *Biochimie.* (2015) 117:3–14. doi: 10.1016/j.biochi.2015.03.014
16. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer.* (2011) 10:38. doi: 10.1186/1476-4598-10-38
17. Harinarayanan J, Reniqua PH, Vamsi KGJ, Alan D, Philip HH, Viswanathan P. The long (lncRNA) and short (miRNA) of it: TGFβ-mediated control of RNA-binding proteins and noncoding RNAs. *Mol Cancer Res*. (2018) 16:567–79. doi: 10.1158/1541-7786.MCR-17-0547
18. Taylor DH, Chu ET, Spektor R, Soloway PD. Long non-coding RNA regulation of reproduction and development. *Mol Reprod Dev*. (2015) 82:932–56. doi: 10.1002/mrd.22581
19. Yang CX, Wang PC, Liu S, Miao JK, Liu XM, Miao YL, et al. Long noncoding RNA 2193 regulates meosis through global epigenetic modification and cytoskeleton organization in pig oocytes. *J Cell Physiol.* (2020) 235:8304–18. doi: 10.1002/jcp.29675
20. Mao HG, Xu XL, Cao HY, Dong XY, Zou XT, Xu NY, et al. Comparative transcriptome profiling of mRNA and lncRNA of ovaries in high and low egg production performance in domestic pigeons (Columba livia). *Front Genet.* (2021) 12:571325. doi: 10.3389/fgene.2021.571325
21. Zhang SH, Kang ZH, Sun XM, Cao XK, Pan CY, Deng RH, et al. Novel lncRNA lncFAM200B: molecular characteristics and effects of genetic variants on promoter activity and cattle body measurement traits. *Front Genet*. (2019). 10:968–80. doi: 10.3389/fgene.2019.00968
22. Miao XY, Luo QM, Zhao HJ, Qin XY. Ovarian transcriptomic study reveals the differential regulation of miRNAs and lncRNAs related to fecundity in different sheep. *Sci Rep.* (2016) 6:35299. doi: 10.1038/srep35299
23. Kim D, Langmead B, Salzberg SLHISAT. a fast spliced aligner with low memory requirements. *Nat Methods.* (2015) 12:357–60. doi: 10.1038/nmeth.3317
24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* (2012) 9:357–9. doi: 10.1038/nmeth.1923
25. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* (2015) 33:243–6. doi: 10.1038/nbt.3172
26. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290–5. doi: 10.1038/nbt.3122

27. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. (2013) 41:e166. doi: 10.1093/nar/gkt646
28. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* (2007) 35:W345–9. doi: 10.1093/nar/gkm391
29. Wang Y, Gao L, Li J, Zhu B, Zhu H, Luo Y, et al. Analysis of long-non-coding RNAs associated with ethylene in tomato. *Gene*. (2018) 674:151–60. doi: 10.1016/j.gene.2018.06.089
30. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* (2005) 21:3674–6. doi: 10.1093/bioinformatics/bti610
31. Schneider JF, Nonneman DJ, Wiedmann RT, Vallet JL, Rohrer GA. Genomewide association and identification of candidate genes for ovulation rate in swine. *J Anim Sci.* (2014) 92:3792–803. doi: 10.2527/jas.2014-7788
32. Nakagawa S, Shimada M, Yanaka K, Mito M, Arai T, Takahashi E, et al. The lncRNA neat is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development.* (2014) 141:4618–27. doi: 10.1242/dev.110544
33. Li Y, Zhang J, Qian Y, Meng C, Wang H, Zhong J, et al. A T > G mutation in the NR5A2 gene is associated with litter size in Hu Sheep through upregulation of promoter activity by transcription factor MTF-1. *Front Genet*. (2019) 10:1011. doi: 10.3389/fgene.2019.01011
34. Quiroz A, Molina P, Santander N, Gallardo D, Rigotti A, Busso D. Ovarian cholesterol efflux: ATP-binding cassette transporters and follicular fluid HDL regulate cholesterol content in mouse oocytes. *Biol Reprod.* (2020) 102:348–61. doi: 10.1093/biolre/ioz159
35. Cesari E, Loiarro M, Naro C, Pieraccioli M, Farini D, Pellegrini L. Combinatorial control of Spo11 alternative splicing by modulation of RNA polymerase II dynamics and splicing factor recruitment during meiosis. *Cell Death Dis*. (2010) 11:240–51. doi: 10.1038/s41419-020-2443-y
36. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* (2009) 136:629–41. doi: 10.1016/j.cell.2009.02.006
37. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature.* (2016) 539:452–5. doi: 10.1038/nature20149
38. Hinshelwood MM, Repa JJ, Shelton JM, Richardson JA, Mangelsdorf DJ, Mendelson CR. Expression of LRH-1 and SF-1 in the mouse ovary: Localization in different cell types correlates with differing function. *Mol Cell Endocrinol.* (2003) 207:39–45. doi: 10.1016/S0303-7207(03)00257-0
39. Meinsohn MC, Hughes CHK, Estienne A, Saatcioglu HD, Pepin D, Duggavathi R, et al. A role for orphan nuclear receptor liver receptor homolog-1 (LRH-1, NR5A2) in primordial follicle activation. *Sci Rep*. (2021) 11:1079. doi: 10.1038/s41598-020-80178-4
40. Schoonjans K, Annicotte JS, Huby T, Botrugno OA, Fayard E, Ueda Y. et a1. Liver receptor homolog 1 controls the expression of the scavenger receptor class B type I. *EMBO Rep*. (2002) 3:1181–7. doi: 10.1093/embo-reports/kvf238
41. Falender AE, Lanz R, Malenfant D, Belanger L, Richards JS. Differential expression of steroidogenic factor-1 and FTF/LRH-1 in the rodent ovary. *Endocrinology.* (2003) 144:3598–610. doi: 10.1210/en.2002-0137
42. Duggavathi R, Volle DH, Mataki C, Antal MC, Messaddeq N, Auwerx J, et al. Liver receptor homolog 1 is essential for ovulation. *Gene Dev*. (2008) 22:1871–6. doi: 10.1101/gad.472008
43. Guo RH, Chen F, Shi ZD. Suppression of notch signaling stimulates progesterone synthesis by enhancing the expression of NR5A2 and NR2F2 in porcine granulosa cells. *Genes*. (2020) 11:120. doi: 10.3390/genes11020120

Conflict of Interest: LC, RB, and SW were employed by Ningbo Sansheng Biological Technology Co., Ltd.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fgene.2022.818810

Transcriptome Sequencing-Based Mining of Genes Associated With Pubertal Initiation in Dolang Sheep

*Zhishuai Zhang, Zhiyuan Sui, Jihu Zhang, Qingjin Li, Yongjie Zhang and Feng Xing**

Key Laboratory of Tarim Animal Husbandry Science and Technology, Xinjang Production and Construction Group, College of Animal Science, Tarim University, Alaer, China

Improving the fertility of sheep is an important goal in sheep breeding as it greatly increases the productivity. Dolang sheep is a typical representative breed of lamb in Xinjiang and is the main local sheep breed and meat source in the region. To explore the genes associated with the initiation of puberty in Dolang sheep, the hypothalamic tissues of Dolang sheep prepubertal, pubertal, and postpubertal periods were collected for RNA-seq analysis on the Illumina platform, generating 64.08 Gb clean reads. A total of 575, 166, and 648 differentially expressed genes (DEGs) were detected in prepuberty_vs._puberty, postpuberty_vs._prepuberty, and postpuberty_vs._puberty analyses, respectively. Based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, the related genes involved in the initiation of puberty in Dolang sheep were mined. Ten genes that have direct or indirect functions in the initiation of puberty in Dolang sheep were screened using the GO and KEGG results. Additionally, quantitative real-time PCR was used to verify the reliability of the RNA-Seq data. This study provided a new approach for revealing the mechanism of puberty initiation in sheep and provided a theoretical basis and candidate genes for the breeding of early-pubertal sheep by molecular techniques, and at the same time, it is also beneficial for the protection, development, and utilization of the fine genetic resources of Xinjiang local sheep.

Keywords: Dolang sheep, RNA-seq, puberty initiation, hypothalamus, differential gene expression

OPEN ACCESS

Edited by:
Yang Zhou,
Huazhong Agricultural University, China

Reviewed by:
Sayed Haidar Abbas Raza,
Northwest A&F University, China
Zhibin Ji,
Shandong Agricultural University, China

****Correspondence:***
Feng Xing
457245195@qq.com

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

Received: *20 November 2021*
Accepted: *26 January 2022*
Published: *03 March 2022*

Citation:
Zhang Z, Sui Z, Zhang J, Li Q, Zhang Y and Xing F (2022) Transcriptome Sequencing-Based Mining of Genes Associated With Pubertal Initiation in Dolang Sheep. Front. Genet. 13:818810. doi: 10.3389/fgene.2022.818810

1 INTRODUCTION

Puberty refers to the age at which an animal first appears to ovulate in heat, indicating an ability to reproduce (Xing et al., 2019). Puberty is influenced by a variety of factors, including genetic mechanisms, nutritional levels, and light hours (Ortavant et al., 1985; Suttie et al., 1985; Greives et al., 2007). The gonadostat hypothesis suggests that as the body develops, hypothalamic *GnRH* neurons become less sensitive to the negative feedback effects of steroid hormones, while pulsatile *GnRH* secretion increases, stimulating gonadotropin secretion and ultimately leading to follicle development and ovulation (Day & Anderson 1998).

Puberty is a physiological phenomenon caused by the development of follicles in the ovaries, and it is regulated by the hypothalamic–pituitary–ovarian axis. Normal or disturbed pubertal development is mainly determined by genetic factors (Roth and Ojeda 2005; Gajdos et al., 2010). Several prior studies have shown that important hypothalamic regulatory gene systems in the initiation of puberty include the leptin system (Hausman et al., 2012), the neurohormone B system (Tusset et al., 2012), the γ-aminobutyric acid system (Terasawa & Fernandez 2001), the Lin28 system (Tommiska et al., 2010), and the microRNA system.

Dolang sheep is the most common sheep breed in the southern region of Xinjiang, known for their high litter rate, adaptability, and perennial heat (Chang et al., 2020). The regulatory mechanism of puberty in Dolang sheep is still unclear, and gene regulation in this regard has not been sufficiently studied in Dolang sheep. Transcriptome sequencing analysis is widely used to screen differential genes, identify candidate genes, analyze metabolic pathways, and predict the relationship between genes and target organs (Marguerat and Bahler 2010; Chen et al., 2011; Ramayo-Caldas et al., 2012). Gao et al. discovered the candidate lncRNA *XLOC_446,331*, which may play a crucial role in regulating female puberty by transcriptome sequencing (Gao et al., 2018). Similarly, Li et al. (2021) identified six (*ESR1*, *NF1*, *APP*, *ENPP2*, *ARNT*, and *DICER1*) genes associated with proestrus by performing differential RNA-seq analysis of hypothalamic tissue in sows' prepuberty, during puberty, and postpuberty. Ling et al. (2014) identified 12 genes associated with high fecundity in Anhui white goats. Based on metabolomics, Zhou et al. (2021) identified polymorphisms in IRS1 were associated with growth efficiency traits in Chinese black Tibetan sheep and Zhang et al. found stall-feeding modified the content of protein and fat, tenderness, water holding capacity, and texture of the longissimus lumborum of Tibetan sheep (Zhang et al., 2021). Based on 16S rRNA gene pyrosequencing, Gui et al. (2021) found that supplementation of concentrate in the cold season improved the rumen microbial abundance of Tibetan sheep. Thus, the biological processes of an organism can be studied in depth using transcriptomics (Conrad et al., 2018). We used transcriptomic and bioinformatics analyses to identify differentially expressed genes in the hypothalamus of Dolang sheep during different periods of puberty and tried to identify candidate genes that might be associated with puberty in Dolang sheep.

2 MATERIALS AND METHODS

2.1 Materials and Treatments

Dolang sheep maintained in the Tarim University experimental station were used as the model in this study. The ewes were observed at 10 o 'clock, 14 o 'clock, and 18 o 'clock every day. The criteria for judging puberty in ewes were mental restlessness, the tendency to walk, the acceptance of ram riding, and the presence of mucus in the vulva. All sheep were in good health. Hypothalamus samples were collected from ewes that were first found to be undergoing puberty, immediately frozen in liquid nitrogen, and stored at −80°C for further analysis. The hypothalami were further collected from prepubertal and postpubertal ewes maintained in the same manner. A total of nine sheep's hypothalamic tissues were collected. Three biological replicates were performed for each period and analyzed.

2.2 Nucleic Acid Extraction and RNA-Seq Library Construction

The total RNA content from the hypothalami was extracted using TRIzol reagent (Beijing Kangwei Century Biotechnology Co.). The purity, concentration, and integrity of the RNA samples were tested by Nanodrop (Thermo Fisher, United States) and Agilent 2100 (Agilent Technologies, United States). A total of 1 μg RNA per sample was used as the input material for the RNA sample preparations. mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using a random hexamer primer and M-MuLV reverse transcriptase. Second strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. Remaining overhangs were converted into blunt ends *via* exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext adapter with a hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 240 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, United States). Then, 3 μl USER enzyme (NEB, United States) was used with size-selected, adapter-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then, PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and index (X) primer. At last, PCR products were purified (AMPure XP system), and library quality was assessed using the Agilent Bioanalyzer 2100 system. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform (Biomarker Technologies), and paired-end reads were generated.

2.3 Sequencing Data Analysis

In the present study, transcript libraries from Dolang sheep at the developmental stages of prepuberty, puberty, and postpuberty were constructed and assayed by high-throughput RNA-seq. Clean reads were obtained using the program Trimmomatic v0.32 (Bolger et al., 2014) by removing reads containing adapters, reads containing poly-N, and low-quality raw reads. Clean reads were then aligned and mapped to the sheep genome (Oar_v4.0) using HISAT2 (Kim et al., 2015). The matched reads were assembled and gene or transcript expression was calculated using StringTie (Pertea et al., 2015). Gene or transcript expression levels were quantified using fragments per kilobase of transcript per million fragments mapped (FPKM) (Florea et al., 2013).

2.4 Sample Correlation Analysis

To assess the reliability of the tested samples, the degrees of variation among the three groups were analyzed using replicate scatter. It was finished by R package corrplot (Wei et al., 2017) and ggplot2 (Wickham 2011).

2.5 Identification and Functional Enrichment Analysis of DEGs

Relative gene expression levels among prepuberty, puberty, and postpuberty were counted using the log2 ratio. The

FIGURE 1 | Ewes undergoing puberty were defined as those receptive to climbing by rams and mucus-laden vulva.

differentially expressed genes (DEGs) were identified using the DEseq2 (Love et al., 2014) method, with the cutoff set as a fold change ≥1.5 and p-value ≤ 0.05. The number of differentially expressed genes among the three groups was compared. Hierarchical clustering analysis was performed on the screened differentially expressed genes using R package heatmap.2 (Warnes et al., 2016), and the genes with the same or similar expression patterns were clustered.

Gene function was annotated *via* alignment with multiple databases, including the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

2.6 Soft Cluster Analysis of DEGs

The soft clustering Mfuzzy function is based on the fuzzy k-means algorithm in e1071 package. R package Mfuzz was used to assign a gene to several clusters using soft clustering methods (Kumar and Futschik 2007). This analysis method can identify potential time series patterns of expression profiles and cluster genes with similar patterns. This can reveal the dynamic patterns of genes and how they are functionally connected.

2.7 qRT-PCR Validation of RNA-Seq Data

Thirteen genes were selected to verify the RNA-Seq results using quantitative real-time PCR (qRT-PCR). Reverse transcription was performed using the reverse transcription kit (TaKaRa). The 15 μl PCR reaction mixture consisted of 5.5 μl ddH_2O, 7.5 μl PerfectStart Green qPCR SuperMix (2×), 1 μl cDNA, and 0.5 μl of each primer (10 μM). The thermal cycle parameters were as follows: 95°C for 15 s, 95°C for 15 s, 55°C for 15 s, and 68°C for 20 s for 40 cycles. Three technical replicates were performed for each sample. Actin was used as an internal reference. The relative expression levels of genes were calculated using the $2^{-\triangle\triangle CT}$ method (Ju et al., 2020). SPSS 17.0 software package (SPSS, Chicago, IL, United States) was applied to analyze the qRT-PCR data.

2.8 Ethical Approval

This study was conducted in accordance with the specifications of the Ethics Committee of the Tarim University of Science and Technology.

3 RESULTS

3.1 Identification of Puberty in Dolang Sheep

Pubertal features of Dolang sheep were observed (**Figure 1**). At the prepuberty stage of ewe, they are calm and do not accept ram for riding, and the vulva is dry. At the pubertal stage, ewes accept ram riding and a moist external pudenda can be observed. At the end of puberty, the ewes become calm, again refused the climbing of ram, and the vulva becomes dry.

3.2 Sequencing Data Quality Control and Sequence Alignment With Reference Genomes

As shown in **Table 1**, after removing reads containing adapters, reads containing ploy-N, and low-quality reads from raw reads, 64.08 GB clean reads and the percentage of Q30 base of each sample was > 92.05%, with an average GC content of 50.79%. Reads were submitted to the NCBI Sequence Read Archive under the accession number PRJNA773843.

The clean reads in each library were then aligned to the sheep genome (Oar_v4.0). The proportion of clean reads mapped to the reference genome ranged from 84.60 to 89.71%, among which 88.74% were uniquely mapped. The percentage of reads mapped to sense strands ranged from 35.09 to 38.63% and those mapped to the antisense strands ranged from 39.35 to 41.95%.

3.3 Quantification of Gene Expression Levels and Correlation Assessment of Biological Replicates

Gene expression levels were estimated by fragments per kilobase of transcript per million fragments mapped. $$FPKM = {cDNA fragments \over { mapped fragments (million) × transcript length (kb)}}$$. We used StringTie software to evaluate gene expression levels (**Figure 2**).

Biological repeat correlation was calculated using R package of the corrplot (**Figure 3**). In the prepuberty, puberty, and postpuberty groups, correlation coefficients among samples were greater than 0.93, 0.85, and 0.95, respectively. These

TABLE 1 | Summary of read numbers of in the prepuberty, puberty, and postpuberty groups.

Sample name	Prepuberty_1	Prepuberty_2	Prepuberty_3	Puberty_1	Puberty_2	Puberty_3	Postpuberty_1	Postpuberty_2	Postpuberty_3
Total clean reads	47,762,312	48,092,774	50,578,284	50,126,846	42,344,556	46,988,672	50,369,858	43,405,702	48,303,884
Clean read q20 (%)	97.14	97.05	97.1	97.04	96.64	97.04	97.09	96.97	97.1
Clean read q30 (%)	92.92	92.73	92.84	92.72	92.05	92.71	92.82	92.57	92.80
Total mapping genome reads	42,408,942	42,929,428	44,954,682	44,971,014	35,823,567	41,661,795	44,724,205	38,712,157	42,397,509
Total mapping genome ratio (%)	88.79	89.26	88.88	89.71	84.60	88.66	88.79	89.19	87.77

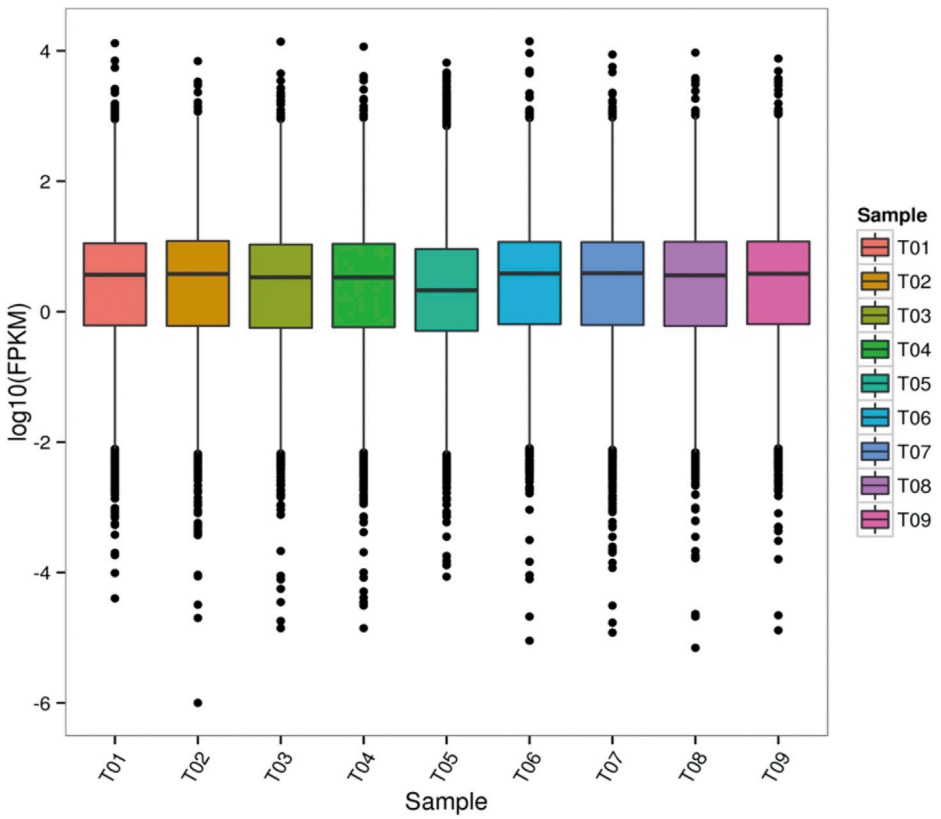


FIGURE 2 | FPKM box line diagram for each sample. T01, T02, and T03 were the prepubertal groups. T04, T05, and T06 were the pubertal groups. T07, T08, and T09 were the postpubertal groups.

results suggest that the sampling of hypothalamus in the present experiment is reliable and suitable for further analysis.

3.4 Identification of Differentially Expressed Genes Among Prepuberty, Puberty, and Postpuberty Ewes

Based on the FPKM method, the transcript abundance of each gene from prepuberty, puberty, and postpuberty data was analyzed (**Figures 4**, **5**). In the comparison of prepuberty and puberty ewes, 575 genes exhibited a significant difference in their expression levels with a threshold of p-value $\leq$ 0.05 and fold change $\geq$ 1.5, including 490 upregulated and 85 downregulated genes. In the comparison of prepuberty and postpuberty, 166 genes exhibited a significant difference in their expression levels with the threshold of p-value $\leq$ 0.05, and fold change $\geq$ 1.5, including 96 upregulated and 70 downregulated genes. In the comparison of puberty and postpuberty, 648 genes exhibited a significant difference in their expression levels

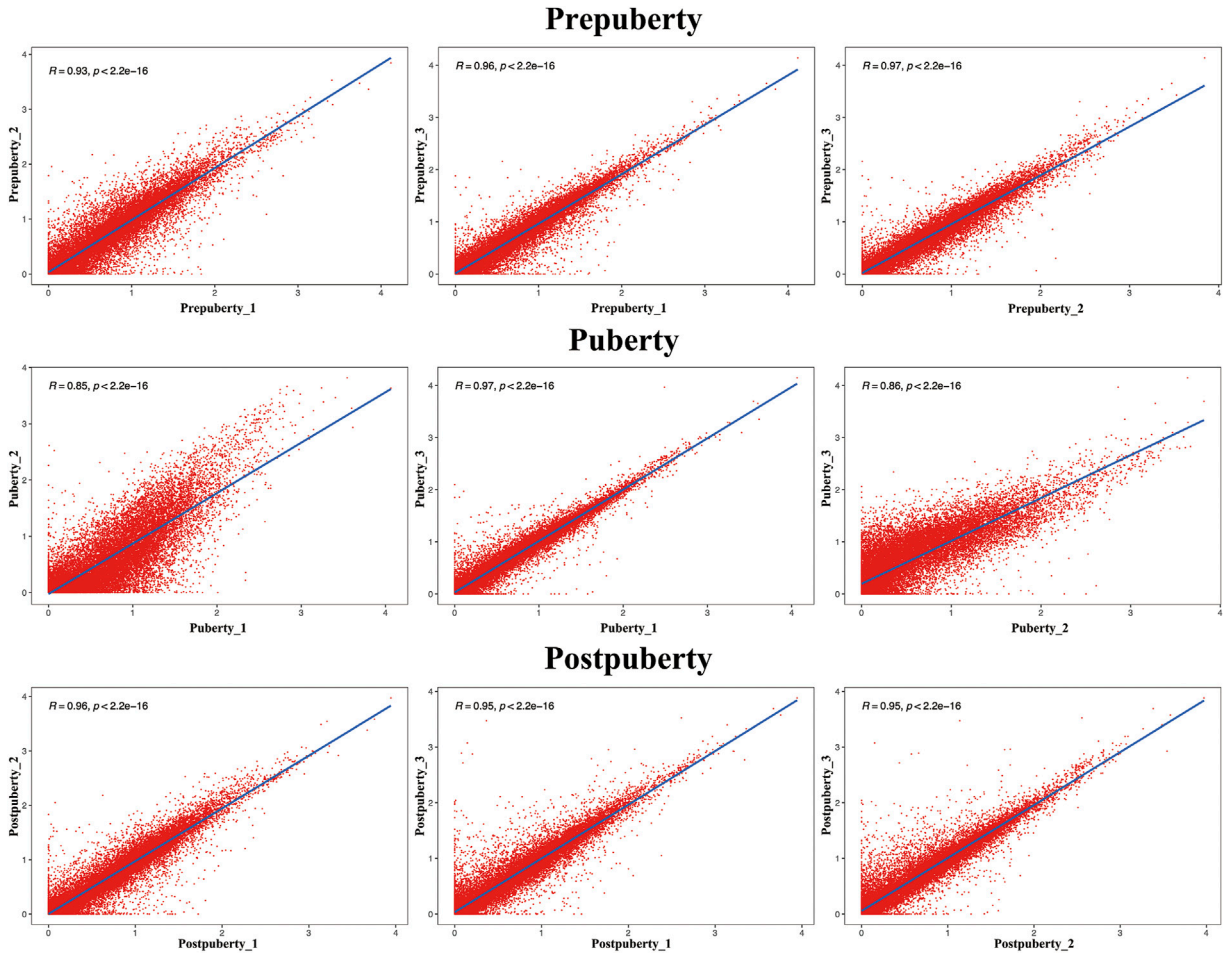


FIGURE 3 | Assessment of correlations among the three groups by analysis of replicate scatter.

with the threshold of p-value ≤ 0.05 and $\log_2$ (fold change) ≥ 1.5, including 97 upregulated and 551 downregulated genes (**Table 2**).

3.5 Functional Enrichment Analysis of DEGs

3.5.1 DEGs Between Prepuberty and Puberty

These identified DEGs were annotated with 20 biological processes, 15 cellular components, and 14 molecular functions in the GO categories (**Figure 6**). The binding and catalytic activities were the top two terms in the molecular function category. In the cellular component category, DEGs were mainly distributed in terms of cell, cell part, and organelle. The most abundant terms in the biological process category were cellular processes and single-organism processes. Reproduction and reproductive processes ranked 13th and 14th, respectively, in the biological process. KEGG pathway analysis classified the DEGs into 199 metabolic pathways.

GO terms associated with puberty were found, including "response to estrogen" (GO: 0043627), "cellular response to gonadotropin stimulus" (GO: 0071371), "copulation" (GO: 0007620), "developmental process involved in reproduction" (GO: 0003006), "female pregnancy" (GO: 0007565), "estrogen receptor binding" (GO: 0030331), and "ovarian follicle development" (GO: 0001541). Some pathways related to puberty were made out, exempli gratia, "estrogen-signaling pathway" (ko04915), "oxytocin-signaling pathway" (ko04921), "GnRH-signaling pathway" (ko04912), and "progesterone-mediated oocyte maturation" (ko04914).

During the progression of prepuberty to puberty, *StAR* expression is upregulated and promotes cholesterol metabolism to pregnenolone in the ovarian steroidogenesis pathway. *GIRK* (G protein-gated inwardly rectifying potassium) showed higher expression levels in the estrogen-signaling pathway in prepuberty than in puberty. Upregulation of the *SOHLH1* and *GAMT* genes was found in ovarian follicle development and reproduction, respectively. *APC/C* (anaphase-promoting complex/cyclosome), *MYT1* (myelin transcription factor 1), and *MAPK* (mitogen-activated protein kinase) were upregulated during progesterone-mediated oocyte maturation. The prolactin signaling pathway members *P38* and *IRF-1* (interferon regulatory factor-1) were upregulated and downregulated, respectively.

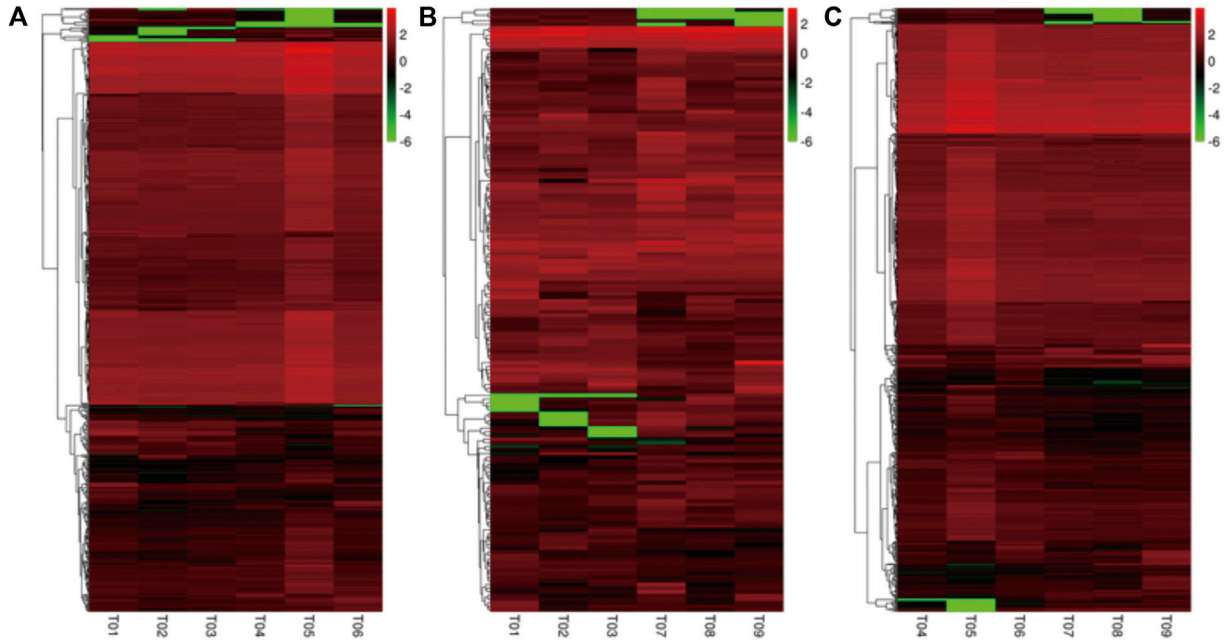


FIGURE 4 | Cluster analysis of all differentially expressed genes. **(A)** Differentially expressed genes between the prepuberty (T01, T02, and T03) and puberty (T04, T05, and T06) groups. **(B)** Differentially expressed genes between the prepuberty (T01, T02, and T03) and postpuberty (T07, T08, and T09) groups. **(C)** Differentially expressed genes between the puberty (T04, T05, and T06) and postpuberty (T07, T08, and T09) groups.

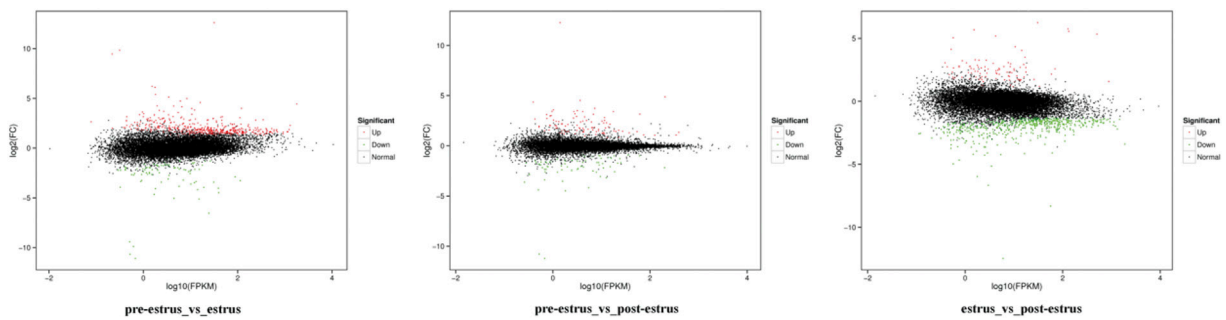


FIGURE 5 | Expression profiles of the identified DEGs. Red and green points represent significant DEGs with p-value ≤ 0.05, and fold change ≥1.5, and black points represent those without significance.

TABLE 2 | The number distribution of differentially expressed genes in different sample group.

DEG set	DEG number	Upregulated	Downregulated
Prepuberty_vs._Puberty	575	490	85
Prepuberty_vs._Postpuberty	166	96	70
Puberty_vs._Postpuberty	648	97	551

3.5.2 DEGs Between Prepuberty and Postpuberty

Identified DEGs were annotated with 20 biological processes, 14 cellular components, and 10 molecular functions in the GO categories (**Figure 7**). The binding and catalytic activities were the top two terms in the molecular function category. In the cellular component category, DEGs were mainly distributed in terms of cell, cell part, and organelle. The most abundant terms in the biological process category were cellular processes and single-organism processes. Reproduction and reproductive processes ranked 13th and 14th in the biological process, respectively. The KEGG pathway analysis classified the DEGs into 94 metabolic pathways.

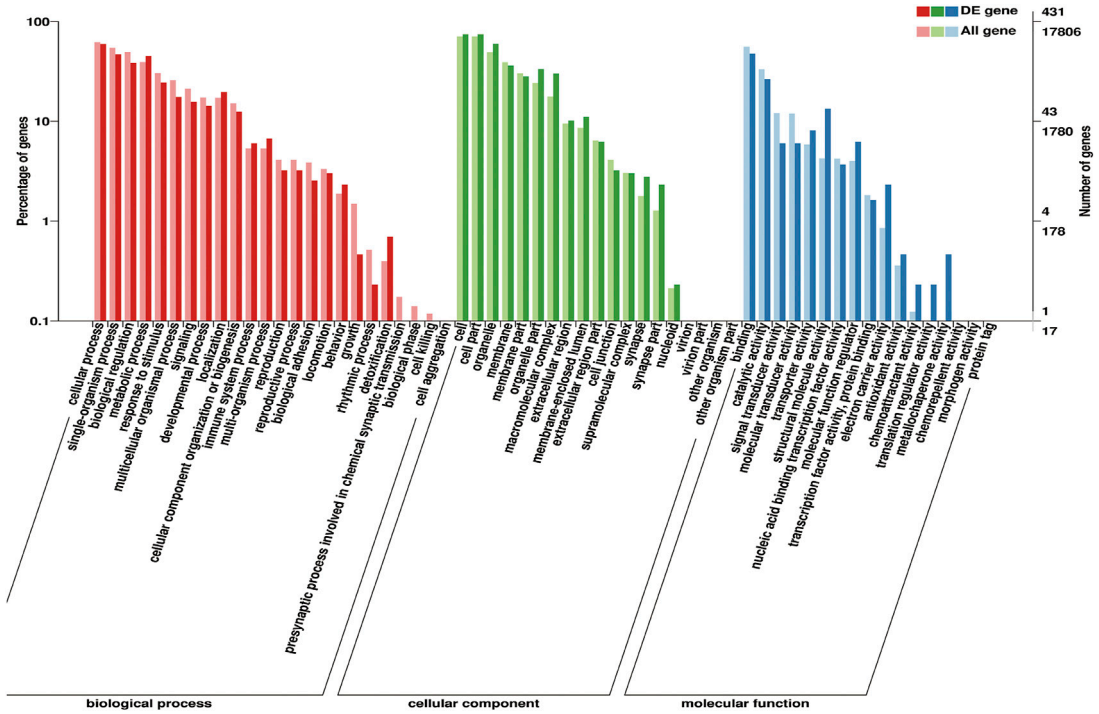


FIGURE 6 | Gene Ontology (GO) classification of 575 DEGs. GO terms are summarized in three main categories: cellular components, molecular functions, and biological processes.

GO terms associated with puberty including "copulation" (GO: 0007620), "ovarian follicle development" (GO: 0001541), "mating behavior" (GO: 0007617), "female pregnancy" (GO: 0007565), "fertilization" (GO: 0009566), "oocyte maturation" (GO: 0001556), and "cell differentiation involved in embryonic placenta development" (GO: 0060706) were included. Several pathways related to puberty, "oxytocin-signaling pathway" (ko04921) and "prolactin-signaling pathway" (ko04917) were further identified as differentially regulated.

As ewes progress from puberty to postpuberty, downregulation of *p38MAPK* leads to a reduction in the extent of its effect on gonadotropin gene expression and secretion. Upregulation of the *DMC1* gene has been observed in ovarian follicle development. *StAR* was downregulated and returned to the same level as in prepuberty ewes. In the oxytocin-signaling pathway, protein kinase C (PKC) and *GIRK* are upregulated. *p38MAPK* (p38 mitogen-activated protein kinase) is downregulated in the GnRH-signaling pathway. In the prolactin-signaling pathway, *p38* was downregulated, and *TH* and *IRF-1* (interferon regulatory factor 1) were upregulated. In progesterone-mediated oocyte maturation, *MAPK*, *Myt1*, and *APC/C* were downregulated.

3.5.3 DEGs Between Puberty and Postpuberty

These identified DEGs were annotated in 20 biological processes, 17 cellular components, and 13 molecular functions in the GO categories (**Figure 8**). The binding and catalytic activities were the top two terms in the molecular function category. In the cellular component category, DEGs were mainly distributed in terms of cell, cell part, and organelle. The most abundant terms in the biological process category were cellular processes and single-organism processes. Reproduction and reproductive processes ranked 13th and 14th, respectively, of the biological process. The KEGG pathway analysis classified the DEGs into 213 metabolic pathways.

GO terms associated with puberty were found, including "post-embryonic development" (GO: 0009791), "developmental process involved in reproduction" (GO: 0003006), "gonad development" (GO: 0008406), "copulation" (GO: 0007620), "fertilization" (GO: 0009566), "mating behavior" (GO: 0007617), "positive regulation of germinal center formation" (GO: 0002636), and "estrogen receptor binding" (GO: 0030331). Some pathways related to puberty were made out, exempli gratia, "ovarian steroidogenesis" (ko04913), "estrogen-signaling pathway" (ko04915), "oxytocin-signaling pathway" (ko04921), "progesterone-mediated oocyte maturation" (ko04914), and "GnRH-signaling pathway" (ko04912).

During the process of ewe development from prepuberty to postpuberty, calmodulin-dependent protein kinase (CaMKK), and soluble guanylyl cyclase (sGC) were found to be up- and downregulated in the oxytocin signaling pathway, respectively. *PRL* and *TH* were upregulated in the PRL signaling pathway. *TTR* was found to be downregulated in hormone activity entry.

3.6 Time Series Expression Clustering Analysis

We selected 811 DEGs that were found to draw clustering maps of the time series (**Figure 9**). By clustering maps of time series and

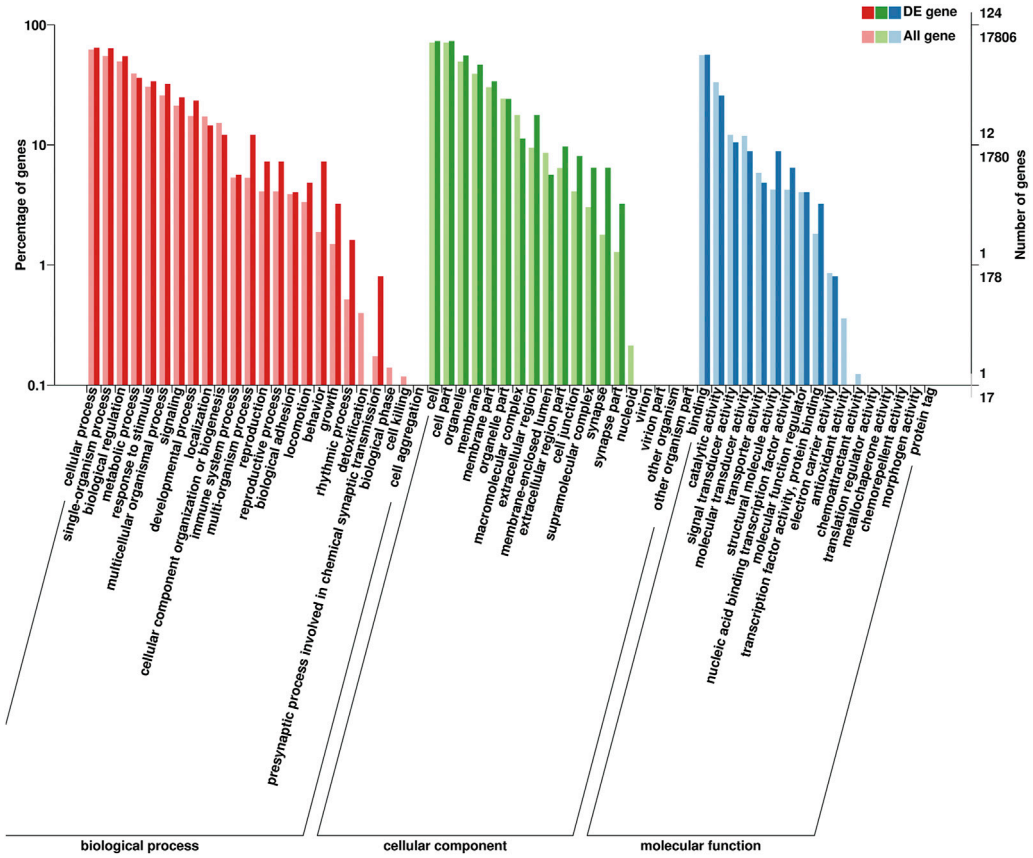


FIGURE 7 | Gene Ontology (GO) classification of 166 DEGs. GO terms are summarized in three main categories: cellular components, molecular functions, and biological processes.

reviewing related literature, we selected the top 10% of genes in each cluster, and identified a portion of genes associated with puberty, for example, the *SPIN1* gene in cluster3, *DMC1* gene in cluster1, *WISP1* genes in cluster2, and *WNT2B* gene in cluster4. All these genes play a direct or indirect role in puberty initiation (**Supplementary Table S1**).

3.7 DEGs Involved in Puberty in Dolang Sheep

Based on these results, a number of genes associated with puberty were identified. The *GAMT* gene plays an important role in embryonic development and the reproductive system (Braissant et al., 2005; Zhao et al., 2021). *SOHLH1* is a multifunctional regulator of the network required for oocyte maintenance and differentiation during early folliculogenesis (Wang et al., 2020). And SOHLH1 is expressed in early oocytes and is necessary for its differentiation (Pangas et al., 2006). *DMC1* is associated with meiosis (Dalman et al., 2019). Loss of function of DMC1 results in defective meiosis and sterility in many species (Chen et al., 2021). The *MACROD1* gene, also known as *LRP16*, plays a role in estrogen signaling (Han et al., 2007; Meng et al., 2007; Tian et al., 2009). It was identified as an estrogen-responsive gene (Han et al., 2003). Estrogen plays a critical role in female puberty and is also important in many aspects of male puberty (Alonso and Rosenfield 2002). *WNT2B* may be regulated during early pregnancy (Atli et al., 2011). Hatzirodos et al. found in the bovine adult ovary that WNT2B is downregulated in the theca interna of large (9–12 mm) compared to small (3–5 mm) healthy follicles (Hatzirodos et al., 2014). *SPIN1* regulates the meiotic cell cycle by modulating the activation of the spindle assembly checkpoint (Choi et al., 2019). *CRH* may play multiple roles in the human endometrium by modulating different signaling cascades (Karteris et al., 2004). *TTR* may be a candidate gene affecting the difference in lambing number in FecB-free mutant small Tail Han sheep (Zhang 2020). *WISP1* plays an important role in embryonic development and immune-related physiological mechanisms (Wu et al., 2012). Co-expression of *WNT2B* and *WISP1* was found by string (https://string-db.org/cgi/input?sessionId=bmuyn8NiPGtp&input_page_show_search=on).

3.8 Expression Profile Analysis by RT-qPCR

To verify the accuracy of the transcriptome sequencing results, we selected 12 differential genes (three genes from prepuberty_vs._puberty and nine genes from puberty_vs._postpuberty) to verify by qRT-PCR and calculated the logarithm of the differential fold of gene expression (**Figure 10**). The correlation coefficients between the two data points were also calculated. The results showed that the correlation coefficient between the RNA-seq and qRT-PCR results was 0.931 ($p < 0.01$). The

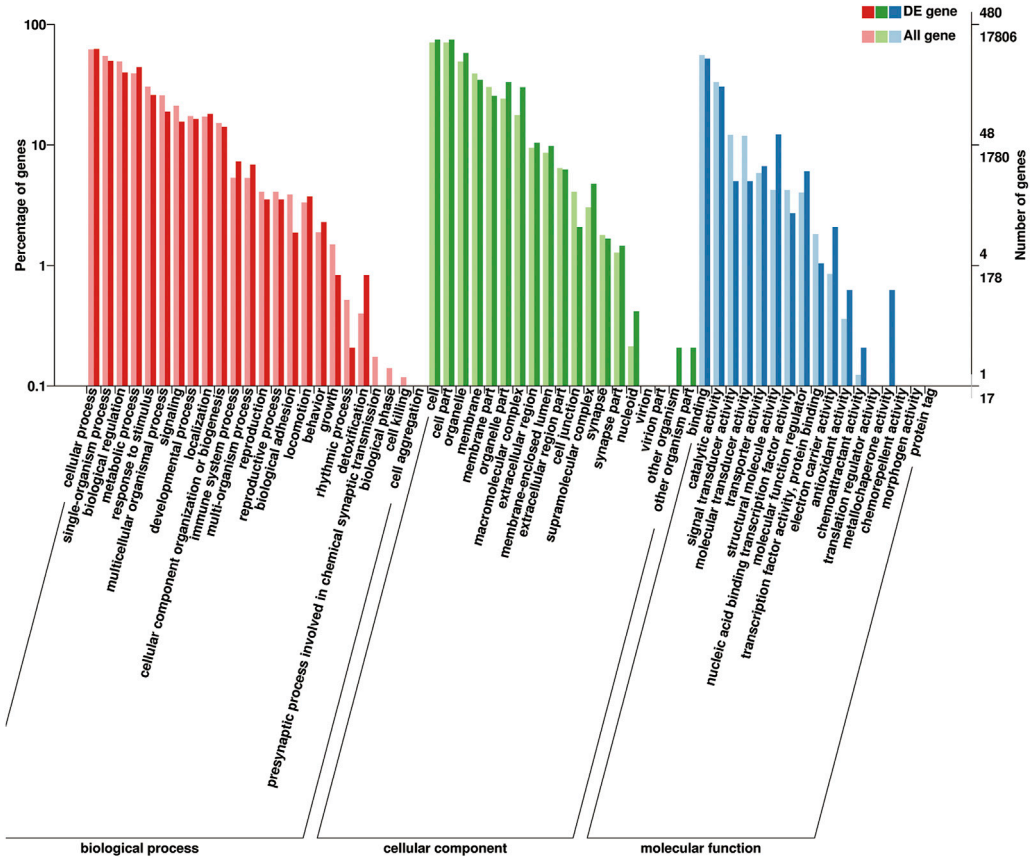


FIGURE 8 | Gene Ontology (GO) classification of 648 DEGs. GO terms are summarized in three main categories: cellular components, molecular functions, and biological processes.

presence of strong correlations indicates that the RNA-seq results were reliable (**Supplementary Table S2**).

4 DISCUSSION

During the initiation of puberty, secondary sexual characteristics begin to develop, the gonads become mature, and ewes gradually acquire the ability to reproduce. In recent years, significant research has been carried out on the initiation of puberty in animals, and many genes associated with the initiation of puberty in sheep have been identified. *NKB* may be an important component of puberty initiation in sheep (Bedenbaugh et al., 2020). Melatonin has a facilitative effect on the initiation of puberty in ewes (Pool et al., 2020). Mutations in the *BMPR-IB* gene cause earlier initiation of puberty in lambs (Wang 2020). *KISS1* expression in the arcuate nucleus increases during puberty in ewes and may be a causative factor in pubertal activation of the reproductive axis. Furthermore, the decrease in *RFRP* expression may be a factor in the initiation of puberty (Li et al., 2020).

In this study, we filtered, assembled, and compared transcriptome data, screened differential genes, and classified genes by bioinformatics analysis, annotated, and functionally classified genes. As such, we identified several genes related to the initiation of puberty in animals. Several software packages, such as HISAT2, StringTie, and DEseq2 were used.

GO classification of Dolang sheep transcriptome suggests that its properties are related to cellular components, biological processes, and molecular functions. Transcriptome analysis of Dolang sheep using the KEGG database identified the estrogen-signaling pathway, ovarian steroidogenesis, oxytocin-signaling pathway, progesterone-mediated oocyte maturation, prolactin-signaling pathway, and GnRH-signaling pathway. These pathways may be associated with the initiation of initial puberty and reproduction in Dolang sheep. Finally, we successfully identified nine genes (*GAMT*, *SOHLH1*, *DMC1*, *MACROD1*, *WNT2B*, *SPIN1*, *CRH*, *TTR*, and *WISP1*) associated with the initiation of sheep primiparity.

The *GAMT* gene plays an important role in embryonic development (Braissant et al., 2005). In this study, the *GAMT* gene was upregulated in prepuberty_vs._puberty and downregulated in puberty_vs._postpuberty. This indicates that the *GAMT* gene starts to initiate expression during puberty in the preparation for embryonic development after mating. *SOHLH1* is a transcriptional regulator that plays a role in the maintenance and survival of primordial ovarian follicles (Bayram et al., 2015). *SOHLH1* is upregulated in prepuberty_vs._puberty, possibly in preparation for sperm–egg cell binding. *DMC1* plays a role in the initiation and progression of meiosis (Habu et al., 1996). In this study, the expression of *DMC1* increased during the initiation and at the end of puberty in Dolang sheep, indicating that *DMC1* may play a role in promoting oocyte meiosis during the initiation of puberty in

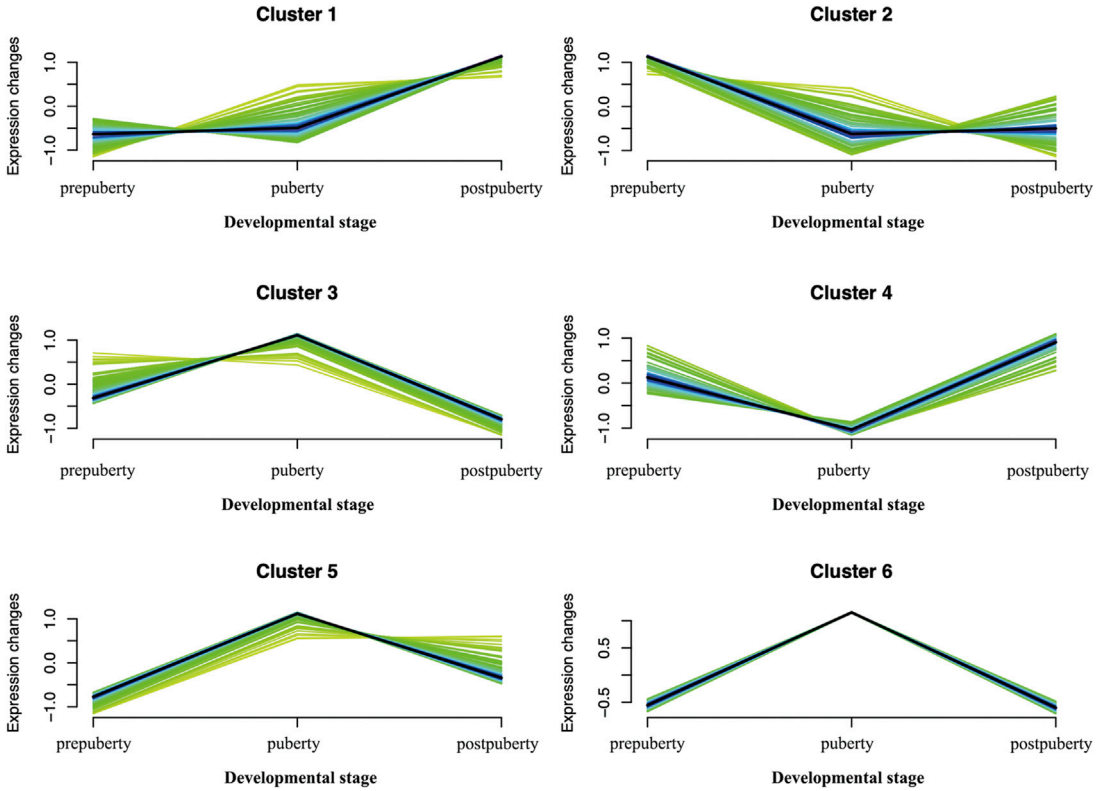


FIGURE 9 | Soft clusters of differently expressed genes. Twelve puberty-related clusters. The horizontal axis represents the different developmental stages (prepuberty, puberty, and postpuberty). The vertical axis represents the changes in expression.

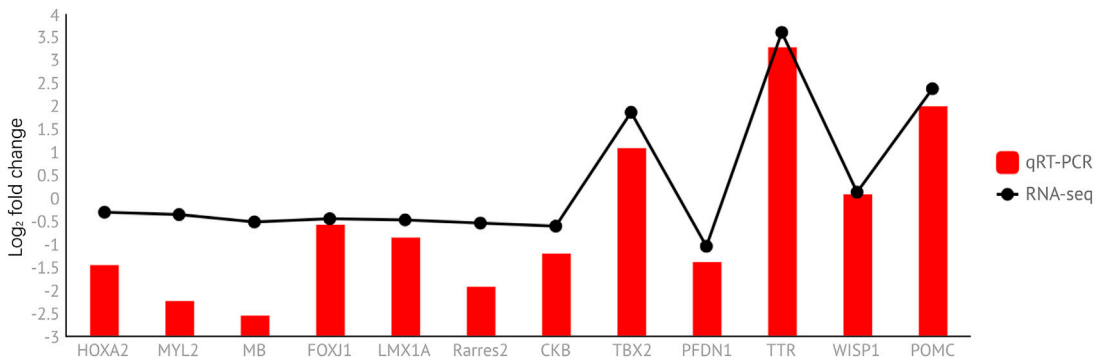


FIGURE 10 | Expression level of DEGs was verified using qRT-PCR and compared with the corresponding data from RNA-Seq assays. The *y*-axis indicates normalized expression levels of the transcripts. The *x*-axis indicates differentially expressed genes.

Dolang sheep. The *MACROD1* gene is also known as *LRP16*, and *LRP16* plays a role in estrogen signaling (Han et al., 2007; Meng et al., 2007; Tian et al., 2009). In prepuberty_vs._puberty, *MACROD1* gene expression is increased, suggesting that *MACROD1* may function during puberty initiation by influencing estrogenic signaling. *WNT2B* and *WISP1* are co-expressed. *Wnt2B* and *WISP1* are enriched in the Wnt-signaling pathway The Wnt-signaling pathway is expressed in granulosa cells, regulated by gonadotropins, and plays a role in follicle development, ovulation, and luteal formation.

5 CONCLUSION

In terms of mRNA expression, *GAMT*, *SOHLH1*, *DMC1*, *MACROD1*, *WNT2B*, *SPIN1*, *CRH*, *TTR*, and *WISP1* genes were significantly different in their expression in the hypothalamus during different pubertal periods in Dolang sheep, suggesting that these genes may be key genes that directly or indirectly influence the initiation of puberty in Dolang sheep. These results provide a basic theoretical basis

for further studies on the molecular mechanisms of pubertal initiation in Dolang sheep.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA773843.

ETHICS STATEMENT

This study was conducted in accordance with the specifications of the Ethics Committee of the Tarim University of Science and Technology.

AUTHOR CONTRIBUTIONS

ZZ conceived the study, conducted the data analysis, and prepared the figures and tables. ZZ, ZS, QL, and JZ performed sample collection and total RNA preparation. ZZ and YZ performed the qRT-PCR validation. ZZ wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant/Award Numbers: 31660652 and 31960655).

ACKNOWLEDGMENTS

We would like to acknowledge Editage (www.editage.com) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.818810/full#supplementary-material

REFERENCES

Alonso, L. C., and Rosenfield, R. L. (2002). Oestrogens and Puberty. *Best Pract. Res. Clin. Endocrinol. Metab.* 16, 13–30. doi:10.1053/beem.2002.0177

Atli, M. O., Guzeloglu, A., and Dinc, D. A. (2011). Expression of Wingless Type (WNT) Genes and Their Antagonists at mRNA Levels in Equine Endometrium during the Estrous Cycle and Early Pregnancy. *Anim. Reprod. Sci.* 125, 94–102. doi:10.1016/j.anireprosci.2011.04.001

Bayram, Y., Gulsuner, S., Guran, T., Abaci, A., Yesil, G., Gulsuner, H. U., et al. (2015). Homozygous Loss-Of-Function Mutations inSOHLH1in Patients with Nonsyndromic Hypergonadotropic Hypogonadism. *J. Clin. Endocrinol. Metab.* 100, E808–E814. doi:10.1210/jc.2015-1150

Bedenbaugh, M. N., Bowdridge, E. C., and Hileman, S. M. (2020). Role of Neurokinin B in Ovine Puberty. *Domest. Anim. Endocrinol.* 73, 106442. doi:10.1016/j.domaniend.2020.106442

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Braissant, O., Henry, H., Villard, A. M., Speer, O., Wallimann, T., and Bachmann, C. (2005). Creatine Synthesis and Transport during Rat Embryogenesis: Spatiotemporal Expression of AGAT, GAMT and CT1. *BMC Dev. Biol.* 5, 9. doi:10.1186/1471-213X-5-9

Chang, W. H., Cui, Z. L., and Wang, J. H. (2020). Identification of Potential Disease Biomarkers in the Ovaries of Dolang Sheep from Xinjiang Using Transcriptomics and Bioinformatics Approaches. *Indian J. Anim. Res.* doi:10.18805/ijar.b-1265

Chen, C., Ai, H., Ren, J., Li, W., Li, P., Qiao, R., et al. (2011). A Global View of Porcine Transcriptome in Three Tissues from a Full-Sib Pair with Extreme Phenotypes in Growth and Fat Deposition by Paired-End RNA Sequencing. *BMC Genomics* 12, 448. doi:10.1186/1471-2164-12-448

Chen, H., He, C., Wang, C., Wang, X., Ruan, F., Yan, J., et al. (2021). RAD51 Supports DMC1 by Inhibiting the SMC5/6 Complex during Meiosis. *Plant Cell* 33, 2869–2882. doi:10.1093/plcell/koab136

Choi, J. W., Zhou, W., Nie, Z. W., Niu, Y. J., Shin, K. T., and Cui, X. S. (2019). Spindlin1 Alters the Metaphase to Anaphase Transition in Meiosis I through Regulation of BUB3 Expression in Porcine Oocytes. *J. Cel Physiol* 234, 8963–8974. doi:10.1002/jcp.27566

Conrad, T., Kniemeyer, O., Henkel, S. G., Krüger, T., Mattern, D. J., Valiante, V., et al. (2018). Module-detection Approaches for the Integration of Multilevel Omics Data Highlight the Comprehensive Response of Aspergillus fumigatus to Caspofungin. *BMC Syst. Biol.* 12, 88. doi:10.1186/s12918-018-0620-8

Dalman, A., Totonchi, M., and Rezazadeh Valojerdi, M. (2019). Human Ovarian Theca-Derived Multipotent Stem Cells Have the Potential to Differentiate into Oocyte-like Cells *In Vitro*. *Cell J* 20, 527–536. doi:10.22074/cellj.2019.5651

Day, M. L., and Anderson, L. H. (1998). suppl_3, 1. doi:10.2527/1998.76suppl_31xCurrent Concepts on the Control of Puberty in Cattle*J. Anim. Sci.*

Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of Exon Skipping Events Differentiate Among Splicing Patterns in Sixteen Human Tissues. *F1000Res* 2, 188. doi:10.12688/f1000research.2-188.v1

Gajdos, Z. K., Henderson, K. D., Hirschhorn, J. N., and Palmert, M. R. (2010). Genetic Determinants of Pubertal Timing in the General Population. *Mol. Cel Endocrinol* 324, 21–29. doi:10.1016/j.mce.2010.01.038

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

Greives, T. J., Mason, A. O., Scotti, M. A., Levine, J., Ketterson, E. D., Kriegsfeld, L. J., et al. (2007). Environmental Control of Kisspeptin: Implications for Seasonal Reproduction. *Endocrinology* 148, 1158–1166. doi:10.1210/en.2006-1249

Gui, L.-S., Raza, S. H. A., Ahmed Allam, F. A. E., Zhou, L., Hou, S., Khan, I., et al. (2021). Altered Milk Yield and Rumen Microbial Abundance in Response to Concentrate Supplementation during the Cold Season in Tibetan Sheep. *Electron. J. Biotechnol.* 53, 80–86. doi:10.1016/j.ejbt.2021.07.001

Habu, T., Taki, T., West, A., Nishimune, Y., and Morita, T. (1996). The Mouse and Human Homologs of DMC1, the Yeast Meiosis-specific Homologous Recombination Gene, Have a Common Unique Form of Exon-Skipped Transcript in Meiosis. *Nucleic Acids Res.* 24, 470–477. doi:10.1093/nar/24.3.470

Han, W.-D., Mu, Y.-M., Lu, X.-C., Xu, Z.-M., Li, X.-J., Yu, L., et al. (2003). Up-regulation of LRP16 mRNA by 17beta-Estradiol through Activation of Estrogen Receptor Alpha (ERalpha), but Not ERbeta, and Promotion of Human Breast Cancer MCF-7 Cell Proliferation: a Preliminary Report. *Endocrine-Related Cancer* 10, 217–224. doi:10.1677/erc.0.0100217

Han, W.-D., Zhao, Y.-L., Meng, Y.-G., Zang, L., Wu, Z.-Q., Li, Q., et al. (2007). Estrogenically Regulated LRP16 Interacts with Estrogen Receptor α and

Enhances the Receptor's Transcriptional Activity. *Endocr. Relat. Cancer* 14, 741–753. doi:10.1677/erc-06-0082

Hatzirodos, N., Hummitzsch, K., Irving-Rodgers, H. F., and Rodgers, R. J. (2014). Transcriptome Profiling of the Theca Interna in Transition from Small to Large Antral Ovarian Follicles. *Plos One* 9, e97489. doi:10.1371/journal.pone.0097489

Hausman, G. J., Barb, C. R., and Lents, C. A. (2012). Leptin and Reproductive Function. *Biochimie* 94, 2075–2081. doi:10.1016/j.biochi.2012.02.022

Ju, Y.-l., Yue, X.-f., Min, Z., Wang, X.-h., Fang, Y.-l., and Zhang, J.-x. (2020). VvNAC17, a Novel Stress-Responsive grapevine (Vitis vinifera L.) NAC Transcription Factor, Increases Sensitivity to Abscisic Acid and Enhances Salinity, Freezing, and Drought Tolerance in Transgenic Arabidopsis. *Plant Physiol. Biochem.* 146, 98–111. doi:10.1016/j.plaphy.2019.11.002

Karteris, E., Papadopoulou, N., Grammatopoulos, D., and Hillhouse, E. (2004). Expression and Signalling Characteristics of the Corticotrophin-Releasing Hormone Receptors during the Implantation Phase in the Human Endometrium. *J. Mol. Endocrinol.* 32, 21–32. doi:10.1677/jme.0.0320021

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317

Kumar, L., and Futschik, M. E. (2007). Mfuzz: a Software Package for Soft Clustering of Microarray Data. *Bioinformation* 2, 5–7. doi:10.6026/97320630002005

Li, Q., Smith, J. T., Henry, B., Rao, A., Pereira, A., and Clarke, I. J. (2020). Expression of Genes for Kisspeptin (KISS1), Neurokinin B (TAC3), Prodynorphin (PDYN), and Gonadotropin Inhibitory Hormone (RFRP) across Natural Puberty in Ewes. *Physiol. Rep.* 8, e14399. doi:10.14814/phy2.14399

Li, Q. N., Pan, X. C., Li, N., Gong, W. T., Chen, Y. S., and Yuan, X. L. (2021). Identification of Circular RNAs in Hypothalamus of Gilts during the Onset of Puberty. *Genes* 12, 15. doi:10.3390/genes12010084

Ling, Y.-H., Xiang, H., Li, Y.-S., Liu, Y., Zhang, Y.-H., Zhang, Z.-J., et al. (2014). Exploring Differentially Expressed Genes in the Ovaries of Uniparous and Multiparous Goats Using the RNA-Seq (Quantification) Method. *Gene* 550, 148–153. doi:10.1016/j.gene.2014.08.008

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

Marguerat, S., and Bähler, J. (2010). RNA-seq: from Technology to Biology. *Cell. Mol. Life Sci.* 67, 569–579. doi:10.1007/s00018-009-0180-6

Meng, Y. G., Han, W. D., Zhao, Y. L., Huang, K., Si, Y. L., Wu, Z. Q., et al. (2007). Induction of the LRP16 Gene by Estrogen Promotes the Invasive Growth of Ishikawa Human Endometrial Cancer Cells through the Downregulation of E-Cadherin. *Cell Res* 17, 869–880. doi:10.1038/cr.2007.79

Ortavant, R., Pelletier, J., Ravault, J. P., Thimonier, J., and Volland-Nail, P. (1985). Photoperiod: Main Proximal and Distal Factor of the Circannual Cycle of Reproduction in Farm Mammals. *Oxf Rev. Reprod. Biol.* 7, 305–345. doi:10.1016/0165-0378(85)90031-2

Pangas, S. A., Choi, Y., Ballow, D. J., Zhao, Y., Westphal, H., Matzuk, M. M., et al. (2006). Oogenesis Requires Germ Cell-specific Transcriptional Regulators Sohlh1 and Lhx8. *Proc. Natl. Acad. Sci.* 103, 8090–8095. doi:10.1073/pnas.0601083103

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Pool, K. R., Rickard, J. P., and de Graaf, S. P. (2020). Overcoming Neuroendocrine and Metabolic Barriers to Puberty: the Role of Melatonin in Advancing Puberty in Ewe Lambs. *Domest. Anim. Endocrinol.* 72, 106457. doi:10.1016/j.domaniend.2020.106457

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., et al. (2012). Liver Transcriptome Profile in Pigs with Extreme Phenotypes of Intramuscular Fatty Acid Composition. *BMC Genomics* 13, 547. doi:10.1186/1471-2164-13-547

Roth, C. L., and Ojeda, S. R. (2005). Genes Involved in the Neuroendocrine Control of normal Puberty and Abnormal Puberty of central Origin. *Pediatr. Endocrinol. Rev.* 3, 67–76.

Suttie, J. M., Gluckman, P. D., Butler, J. H., Fennessy, P. F., Corson, I. D., and Laas, F. J. (1985). Insulin-like Growth Factor 1 (IGF-1) Antler-Stimulating Hormone? *Endocrinology* 116, 846–848. doi:10.1210/endo-116-2-846

Terasawa, E., and Fernandez, D. L. (2001). Neurobiological Mechanisms of the Onset of Puberty in Primates. *Endocr. Rev.* 22, 111–151. doi:10.1210/edrv.22.1.0418

Tian, L., Wu, Z., Zhao, Y., Meng, Y., Si, Y., Fu, X., et al. (2009). Differential Induction of LRP16 by Liganded and Unliganded Estrogen Receptor α in SKOV3 Ovarian Carcinoma Cells. *J. Endocrinol.* 202, 167–177. doi:10.1677/joe-09-0054

Tommiska, J., Wehkalampi, K., Vaaralahti, K., Laitinen, E. M., Raivio, T., and Dunkel, L. (2010). LIN28B in Constitutional Delay of Growth and Puberty. *J. Clin. Endocrinol. Metab.* 95, 3063–3066. doi:10.1210/jc.2009-2344

Tusset, C., Noel, S. D., Trarbach, E. B., Silveira, L. F. G., Jorge, A. A. L., Brito, V. N., et al. (2012). Mutational Analysis of TAC3 and TACR3 Genes in Patients with Idiopathic central Pubertal Disorders. *Arq Bras Endocrinol. Metab.* 56, 646–652. doi:10.1590/s0004-27302012000900008

Wang, Y. T. (2020). *Effects of Different BMPR-IB Genotypes on Reproductive Hormones, Growth and Development in Puberty of Sheep*. Shihezi: Shihezi University. (in Chinese).

Wang, Z., Liu, C.-Y., Zhao, Y., and Dean, J. (2020). FIGLA, LHX8 and SOHLH1 Transcription Factor Networks Regulate Mouse Oocyte Growth and Differentiation. *Nucleic Acids Res.* 48, 3525–3541. doi:10.1093/nar/gkaa101

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). *Package 'gplots'. Various R Programming Tools for Plotting Data*.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2017). Package 'corrplot. *Statistician* 56, e24.

Wickham, H. (2011). ggplot2. *Wires Comp. Stat.* 3, 180–185. doi:10.1002/wics.147

Wu, J. J., Li, W. M., Zhao, R. X., Wang, C., and Zhang, S. J. (2012). Polymorphisms in WISP1 and NMI Genes and Their Effects on the Mortality of Chick Embryos. *J. Poult. Sci.* 49, 249–253. doi:10.2141/jpsa.0110159

Xing, F., Zhang, C., and Kong, Z. (2019). Cloning and Expression of Lin-28 Homolog B Gene in the Onset of Puberty in Duolang Sheep. *Asian-australas J. Anim. Sci.* 32, 23–30. doi:10.5713/ajas.18.0276

Zhang, X., Han, L., Hou, S., Raza, S. H. A., Wang, Z., Yang, B., et al. (2022). Effects of Different Feeding Regimes on Muscle Metabolism and its Association with Meat Quality of Tibetan Sheep. *Food Chem.* 374, 131611. doi:10.1016/j.foodchem.2021.131611

Zhang, Z. b. (2020). *Screening Polytocous Candidate Genes in Small Tail Han Sheep Based on Hypothalamic Multi-Omics Analysis*. Beijing: Chinese Academy of Agricultural Sciences Thesis. (in Chinese).

Zhao, X. R., Nie, C. S., Zhang, J. X., Li, X. H., Zhu, T., Guan, Z., et al. (2021). Identification of Candidate Genomic Regions for Chicken Egg Number Traits Based on Genome-wide Association Study. *BMC Genomics* 22, 11. doi:10.1186/s12864-021-07755-3

Zhou, L., Raza, S. H. A., Gao, Z.-h., Sayed, S. M., Shukry, M., Abd El-Aziz, A. H., et al. (2021). Variations in the Insulin Receptor Substrate 1 (IRS1) and its Association with Growth Traits in Chinese Black Tibetan Sheep (*Ovis aries*). *Anim. Biotechnol.* 32, 786–791. doi:10.1080/10495398.2021.1957687

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 in Genetics

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fgene.2022.846345

The Dynamic of PRAMEY Isoforms in Testis and Epididymis Suggests Their Involvement in Spermatozoa Maturation

*Chandlar H. Kern, Weber B. Feitosa† and Wan-Sheng Liu**

Department of Animal Science, Center for Reproductive Biology and Health, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA, United States

The preferentially expressed antigen in melanoma, Y-linked (PRAMEY) is a cancer/testis antigen expressed predominantly in bovine spermatogenic cells, playing an important role in germ cell formation. To better understand PRAMEY's function during spermatogenesis, we studied the dynamics of PRAMEY isoforms by Western blotting (WB) with PRAMEY-specific antibodies. The PRAMEY protein was assessed in the bovine testicular and epididymal spermatozoa, fluid and tissues, and as well as in ejaculated semen. The protein was further examined, at a subcellular level in sperm head and tail, as well as in the subcellular components, including the cytosol, nucleus, membrane, and mitochondria. RNA expression of PRAMEY was also evaluated in testis and epididymal tissues. Our WB results confirmed the previously reported four isoforms of PRAMEY (58, 30, 26, and 13 kDa) in the bovine testis and spermatozoa. We found that testicular spermatozoa expressed the 58 and 30 kDa isoforms. As spermatozoa migrated to the epididymis, they expressed two additional isoforms, 26 and 13 kDa. Similarly, the 58 and 30 kDa isoforms were detected only in the testis fluid, while all four isoforms were detected in fluid from the cauda epididymis. Tissue evaluation indicated a significantly higher expression of the 58 and 13 kDa isoforms in the cauda tissue when compared to both the testis and caput tissue ($p < 0.05$). These results indicated that testis samples (spermatozoa, fluid, and tissue) expressed predominantly the 58 and 30 kDa PRAMEY isoforms, suggesting their involvement in spermatogenesis. In contrast, the 26 kDa isoform was specific to epididymal sperm and the 13 kDa isoform was marked in samples derived from the cauda epididymis, suggesting their involvement in sperm maturation. Results from the sperm head and tail experiments indicated that the 13 kDa isoform increased 4-fold in sperm tails from caput to cauda, suggesting this isoform may have a significant role in tail function. Additionally, the 13 kDa isoform increased significantly ($p < 0.05$) in the cytosol during epididymal passage and tended to increase in other subcellular components. The expression of PRAMEY in the sperm subcellular components during epididymal maturation suggests the involvement of PRAMEY, especially the 13 kDa isoform, in sperm motility.

Keywords: PRAMEY, cancer/testis antigen, spermatogenesis, sperm maturation, cattle

OPEN ACCESS

Edited by:
Aixin Liang,
Huazhong Agricultural University, China

Reviewed by:
Tao Zeng,
Zhejiang Academy of Agricultural Sciences, China
Elisabeth Pinart,
University of Girona, Spain

****Correspondence:***
Wan-Sheng Liu
wul12@psu.edu

†Weber B. Feitosa, Center of Natural and Human Science, Federal University of ABC, Santo Andre, Brazil

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

Received: *31 December 2021*
Accepted: *08 February 2022*
Published: *21 March 2022*

Citation:
Kern CH, Feitosa WB and Liu W-S (2022) The Dynamic of PRAMEY Isoforms in Testis and Epididymis Suggests Their Involvement in Spermatozoa Maturation. Front. Genet. 13:846345. doi: 10.3389/fgene.2022.846345

INTRODUCTION

Spermatozoa are formed throughout a male's reproductive lifetime from spermatogonial stem cells (SSCs) and function as the male reproductive cells that contribute to fertilization. Spermatogenesis takes place in the seminiferous tubules of the testis through three phases: mitosis, meiosis, and spermiogenesis. The last phase, spermiogenesis, represents the process where haploid round spermatids are converted into fully differentiated spermatozoa and are released into the lumen of the seminiferous tubules (Hess and de Franca, 2008).

During spermatogenesis, numerous germ cell-specific antigens are expressed, however many of these antigens are nonexistent or are minimally detected in normal somatic tissues, while they are highly expressed in various tumors (Chang et al., 2011). Because these antigens have restricted expression in the testis and cancer cells, they have been termed cancer/testis antigens (CTAs). During spermatogenesis, numerous CTAs are detected at a specific stage (e.g., synaptonemal complex protein 1, SCP1) (Meuwissen et al., 1992), while others are found at multiple time points during spermatogenesis (e.g. trophinin, TRO, preferentially expressed antigen in melanoma-like 1, PRAMEL1, and Prame family 12, PRAMEF12) (Saburi et al., 2001; Mistry et al., 2013; Wang et al., 2019; Kern et al., 2021). The stage-specific appearance of CTAs throughout the spermatogenic process has led researchers to believe that there is significant importance associated with CTA emergence and their function in spermatogenesis (Tulsiani et al., 1998).

One of the CTAs, the PRAME protein, was originally discovered in a human melanoma cell line (Ikeda et al., 1997). Primary research in cancer biology found PRAME to be a dominant repressor of the retinoic acid receptor (RAR) in melanoma cells (Epping et al., 2005, 2008). Later studies indicated that PRAME was involved in nuclear factor Y (NFY)-mediated transcriptional regulation as a subunit of a Cullin2-based E3 ubiquitin ligase in leukemia cells (Costessi et al., 2011). Members of the *PRAME* gene family encode leucine-rich repeat (LRR) proteins that fold into a horseshoe shape, which provides a versatile structural framework for the formation of protein-protein interactions (Kobe and Kajava, 2001; Epping et al., 2005; Wadelin et al., 2010). Protein phosphatase 1, regulatory (inhibitor) subunit 7 (PPP1R7), also known as SDS22, is another LRR protein and interacts with protein phosphatase 1 catalytic subunit gamma isozyme (PPP1CC), also known as PP1γ2, in bovine cauda epididymal spermatozoa (Huang et al., 2002; Mishra et al., 2003). As a testis/spermatozoa-specific phosphatase, PP1γ2 acts as an important regulator of sperm motility, and male fertility (Smith et al., 1996; Fardilha et al., 2011). While the regulatory function of PRAME in cancer cells is well studied (Epping et al., 2005; Costessi et al., 2011), the role of the PRAME protein family in germ cells, and male reproduction is still unclear.

Through evolution, the *PRAME* gene family has been amplified and it constitutes a large gene family in eutherian mammals (Birtle et al., 2005; Church et al., 2009; Chang et al., 2011). In bovine, *PRAME* consists of multiple copies in chromosome 16, with a single copy in chromosome 17. An autosome (i.e., chromosome 17)-to-Y transposition and subsequent amplification resulted in a Y-linked *PRAME* gene (*PRAMEY*) sub-family (Chang et al., 2011, 2013). This transposition event occurred during bovine evolution and is believed to enhance bovid male fertility (Chang et al., 2011). Additionally, the copy number variation (CNV) of PRAMEY has been found to correlate with male reproductive traits and could possibly be a valuable marker for male fertility selection (Yue et al., 2013).

Previous studies identified four PRAMEY isoforms (58, 30, 26, and 13 kDa) within the bovine testes tissue (58 and 30 kDa isoforms) and/or epididymal spermatozoa (30, 26, and 13 kDa isoforms) using a custom-made PRAMEY-specific antibody (Liu et al., 2017). The bovine PRAMEY was also found to be enriched in the intermitochondrial cement (IMC) and chromatoid body (CB) of bovine spermatogenic cells (Liu et al., 2017). IMC and CB represent a class of mammalian spermatogenic cell-specific organelles known as germinal granules and commonly referred to as nuage (Meikar et al., 2011). The IMC originates in the cytoplasm of spermatocytes, but it can no longer be located after meiosis. However, the CB is present in the cytoplasm of post-meiotic spermatids. Therefore, PRAMEY's presence in these organelles signifies its importance in spermatogenesis.

Given the knowledge of PRAMEY's possible role in spermatogenesis, the objective of this study was to characterize the detailed expression pattern of the PRAMEY protein within spermatozoa, reproductive fluids and tissues of the bovine testis, epididymis, and ejaculated semen. Ultimately, our goal was to use the expression and localization data gained from this study to understand the functional role of PRAMEY during sperm formation, maturation, and function.

MATERIAL AND METHODS

Sample Collection

Pairs of mature bovine testes with intact epididymides were obtained from a local slaughter house (Nicholas Meat, LLC, Loganton, PA, United States). Testes were transported to the laboratory in ice-filled coolers and processed within 2 h to collect spermatozoa, fluid, and tissues from the testes and epididymides (see following sections for specific collection procedures).

For the experiments that evaluated PRAMEY expression in spermatozoa, fluid, and tissue, samples were collected from five pairs of mature testes with intact epididymides (n = 5). For the experiments that evaluated the PRAMEY expression in sperm head and tail, as well in the sperm subcellular fraction, spermatozoa were collected from three pairs of mature testes with intact epididymides (n = 3). In the experiments where mRNA expression was analyzed, the samples (testis, caput, corpus, and cauda epididymis) were collected from three animals (n = 3). For the experiments using freshly ejaculated semen and seminal plasma, the samples were collected from nine yearling beef bulls (n = 9) from the Pennsylvania Department of Agriculture Livestock Evaluation Center.

Testicular and Epididymal Sperm Collection

Epididymides were removed from the testis and dissected into caput and cauda segments. The PRAMEY mRNA expression experiment was the only one to include the corpus epididymal segment. Sperm collection procedures were adapted from protocols previously established in our laboratory (Liu et al., 2017). Briefly, the isolated testicles were washed with phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na2PO4, and 1.8 mM KH2PO4, pH 7.4) and opened by a sagittal cut using a scalpel. By applying light pressure to the testis, the testicular fluid containing spermatozoa was collected, transferred to a 2 ml Eppendorf tube, and centrifuged at 1,500 x g for 10 min at 4 °C to remove the testicular spermatozoa (the supernatant was reserved). The testicular sperm pellet was washed twice in 2 ml PBS at 1,500 x g for 10 min at 4°C and the resulting sperm pellet was stored at −80°C. The dissected caput or cauda epididymides were washed in PBS and placed in a Petri dish. The epididymal tissue was cut with a scalpel and the fluid containing the spermatozoa was collected by applying pressure on the caput or cauda epididymides. The epididymal fluid was transferred to a 2 ml Eppendorf and centrifuged at 1,500 x g for 10 min at 4°C to collect the epididymal spermatozoa (the supernatant was reserved). The caput and cauda sperm pellets were washed twice in 2 ml PBS at 1,500 x g for 10 min at 4°C and the resulting sperm pellets were stored at −80°C. For the sperm head and tail separation portion of the project, the caput or cauda epididymal spermatozoa were isolated by slicing the epididymal tissue in a Petri dish with PBS. The isolated spermatozoa were centrifuged at 1,500 x g for 10 min at 4°C, and the caput and cauda sperm pellets were used to fractionate the spermatozoa into head and tail portions by sonication.

Testis and Epididymal Fluid Collection

After collecting spermatozoa from the bovine testis and epididymis as described above, the reserved supernatants were transferred to new 2 ml Eppendorf tubes and centrifuged at 4,000 x g for 20 min at 4°C to remove cellular debris. The testis, caput and cauda fluids were stored at −80°C.

Testis and Epididymal Tissue Collection

Tissue collection procedures were adapted from protocols previously established in our laboratory (Liu et al., 2017). Tissues from the testis and caput and cauda segments of the epididymis were collected using a scalpel, cut into small pieces, and washed twice in PBS. A portion of the tissues were snap frozen in liquid nitrogen and stored at −80°C until further use for protein and RNA extraction.

Ejaculated Semen Collection

Ejaculated semen was obtained from yearling beef bulls at the Pennsylvania Department of Agriculture Livestock Evaluation Center from animals within the Bull Testing Program. The bulls were assessed through breeding soundness evaluation and electroejaculation was used to obtain semen by a certified veterinarian. All animal procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the Animal Care and Use Committees of the Pennsylvania State University. The collected semen was donated for our research. Semen was collected in 15 ml tubes, kept on ice for the duration of collection, and was transported to the laboratory in ice-filled coolers.

Sperm Head and Tail Separation

The sperm head and tail separation procedure was adapted from Tateno and co-workers (Tateno et al., 2000). The isolated spermatozoa from epididymal fluid were washed three times in PBS (containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na2PO4, and 1.8 mM KH2PO4, pH 7.4) with centrifugation at 1,100 x g for 3 min at 4°C. Each sample was sonicated for 30 s at 8 W of power (Fisher Scientific, Sonic Dismembrator Model 100) to break the sperm heads from the tails. To separate the head and tail fractions of the sonicated mixture, 500 uL of Tris Buffer Saline (TBS) (containing 50 mM Tris-Cl, 150 mM NaCl, pH 7.6) was added to each sperm sample and then mixed by vortexing. The diluted spermatozoa (200 uL) were added to 1.0 ml of a 90% Percoll solution (GE Healthcare, product no. 17-0891-02) and the samples were centrifuged at 15,000 x g for 15 min at room temperature (RT). After centrifugation, the sperm tails were located in a thin band near the top of the 1.5 ml tube, and the sperm heads were located in a pellet at the bottom of the tube. The sperm heads and tails were removed separately from the 90% Percoll gradient and put into new 1.5 ml tubes to wash. The sperm heads and tails were washed in 500 uL of TBS (containing 50 mM Tris-Cl, 150 mM NaCl, pH 7.6) at 9,000 x g for 5 min at RT and were immediately frozen in liquid nitrogen.

Subcellular Fractionation

Subcellular fractionation was performed following the Abcam subcellular fractionation protocol (https://www.abcam.com/protocols/subcellular-fractionation-protocol). Briefly, sperm cells from caput and cauda epididymides were resuspended in 500 µl fractionation buffer and (20 mM HEPES, 10 mM KCL, 2 mM MgCl2, 1 mM EDTA, and 1 MM EGTA supplemented with 1 mM DTT), protease and phosphatase inhibitor cocktails. They were homogenized and incubated for 20 min on ice and were then centrifuged at 720 x g for 5 min. The supernatant containing the sperm cytoplasm, plasma membrane, and mitochondria was transferred into a fresh tube and kept on ice, while the pellet containing nuclei was washed with 500 µl fractionation buffer. The nuclear pellet was dispersed by vortexing, and it was then centrifuged again at 720 x g for 10 min. The resulting pellet was resuspended in TBS with 0.1% SDS, sonicated briefly to shear genomic DNA and homogenize the lysate (3 s on ice at a power setting of 2-continuous), and stored at −80°C. The supernatant containing the sperm cytoplasm, membrane, and mitochondria was centrifuged at 10,000 x g for 5 min at 4°C. The supernatant containing the cytoplasm and membrane was transferred into a fresh tube and kept on ice, while the pellet containing the mitochondria was processed as described for the nuclear pellet to obtain mitochondrial lysate in TBS/0.1% SDS. The supernatant containing the cytoplasm and membrane was centrifuged at 50,000 x g for 2 h. The supernatant containing the cytosol was stored at −80°C, while the pellet was resuspended in 400 µl of

fractionation buffer supplemented with 1% Triton and re-centrifuged for 45 min under the same conditions. The resulting membrane pellet was processed as described for the nuclear pellet.

Protein Extraction

Protein extraction procedures were adapted from protocols previously established in our laboratory (Liu et al., 2017). Protein was extracted from testicular and epididymal spermatozoa, fluids, and tissues, and as well as from sperm heads and tails for this project. Protein was extracted using CelLytic buffer (Sigma, product no. C3228) added with protease inhibitor cocktail (Thermo Scientific, product no. 1860932) and phosphatase inhibitor cocktail (Thermo Scientific, product no. 1862495). The testes and epididymal sperm, fluid, and tissues, and sperm head and tail pellets were removed from the −80°C freezer and ice-cold extraction buffer (supplemented with protease inhibitor and phosphatase inhibitor cocktails) was immediately added to the pellets. The mixture was homogenized on ice and then incubated on ice for 10 min at 4°C on a shaker. The samples were centrifuged for 10 min at 13,200 x g at 4°C. The supernatants were removed and were referred to as the respective protein for each testis and epididymal sperm, fluid, tissue, and sperm head, and tail samples in this study.

RNA Extraction, cDNA Synthesis, and RT-PCR

The following kits were used for RNA extraction from the testis, and the caput, corpus, and cauda epididymal tissues: The miRNAeasy Mini Kit (Qiagen cat no. 217004), QIAshredder spin columns (Qiagen cat no. 79654), and the RNase-free DNase set (Qiagen cat no. 79254). The protocol for RNA extraction was performed following the "Purification of Total RNA, Including Small RNAs, from Animal Tissues" described in the miRNeasy Mini Handbook (file:///C:/Users/wul12/AppData/Local/Temp/HB-1253003_HB_miRNeasy_96_1,120_%20WW.pdf). cDNA was synthesized using the Superscript III First-Strand Synthesis System for reverse transcription (RT)-PCR (Invitrogen product no. 18080051). The cDNA synthesis was performed following the protocol under "First-strand cDNA synthesis" within the product handbook. RT-PCR was performed using the cDNA samples, The PCR protocol was as follows: each 20-µl reaction contained 13.76 µl of distilled water, 0.5 µl of each primer (10 pmol/µl), 4 µL of Bioline 5× buffer (Bioline United States Inc., Taunton, MA, including 200 µM deoxyribonucleotide triphosphates), 0.24 µl of Bioline Taq DNA polymerase (Bioline United States Inc.), and 1 µl of either genomic DNA (10 ng/µl) or water. Thermocycling for the gene of interest, PRAMEY (forward, 5′- TCAGGACCT GGAGGTCAAC-3'; reverse, 5′-TGTGGCAATATGTGGATG CG-3′), consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles of at 94°C for 30 s, at 65°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 5 min. Thermocycling for the control gene, GADPH (forward, 5′- AACGGATTTGGC CGTATTGG-3'; reverse, 5′-CATTCTCGGCCTTGACTGTG-3′), consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles of at 94°C for 30 s, at 55°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 5 min.

Gel Electrophoresis and Western Blotting

Gel electrophoresis and WB procedures were adapted from protocols previously established in our laboratory (Liu et al., 2017). The protein extracts were separated by a 4–12% acrylamide gel. The gels were electronically transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, product no. IPVH00010). The membranes were blocked in 5% nonfat dried milk (NFDM) in tris buffered saline containing 0.05% Tween-20 (TBST). After being briefly washed in TBST, the membrane was incubated in the primary antibody, a PRAMEY-specific (Liu et al., 2017) or PRAMEYc custom antibody produced by New England Peptide, LLC (Gardner, MA, United States) at a dilution of 1:500, overnight at 4°C. The membranes were washed with TBST 3 times for 10 min each and then incubated in the secondary antibody, anti-rabbit IgG, HRP linked (Cell Signaling Technology, product no. 7074S, 1: 5,000 dilution) for 1 h at room temperature. The membranes were washed 3 times for 10 min each and the reactive proteins were detected by SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific, product no. 34095). After PRAMEY or PRAMEYc antibody incubation, membranes were stripped with 15 ml of stripping buffer (10% SDS, 0.5 M Tris HCl, β-mercaptoethanol) for 30 min at 50°C. Membranes were washed with TBST 6 times for 5 min each and then re-blocked in 5% NFDM. Alpha tubulin (TUBA) primary antibody (Cell Signaling Technology, product no. 3873, 1:1,000 dilution) was used to probe the membranes again overnight at 4°C. The membranes were washed with TBST 3 times for 10 min each and then incubated in the secondary antibody, anti-mouse IgG, HRP linked (Cell Signaling Technology, product no. 7076, 1:5,000 dilution) for 1 h at room temperature. The membranes were washed 3 times for 10 min each and the reactive proteins were detected by SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific, product no. 34095).

Statistical Analysis

WB images were analyzed using ImageJ software (Schneider et al., 2012). The data were analyzed by the normality test (Shapiro-Wilk test) and equal variance test (Brown-Forsythe) using SigmaPlot 12.0 (Statistical Software). After meeting the assumptions of normally distributed data and homogeneity of variance, the difference in treatment levels was compared by one-way ANOVA with the post-hoc *Tukey test*. Data are expressed as the mean ± SEM, and a value of $p < 0.05$ was considered statistically significant.

RESULTS

The Pattern of PRAMEY Expression in Bovine Sperm

In a previous report, four different isoforms of the bovine PRAMEY protein including the 58, 30, 26, and 13 kDa isoforms were identified by WB with a PRAMEY-specific

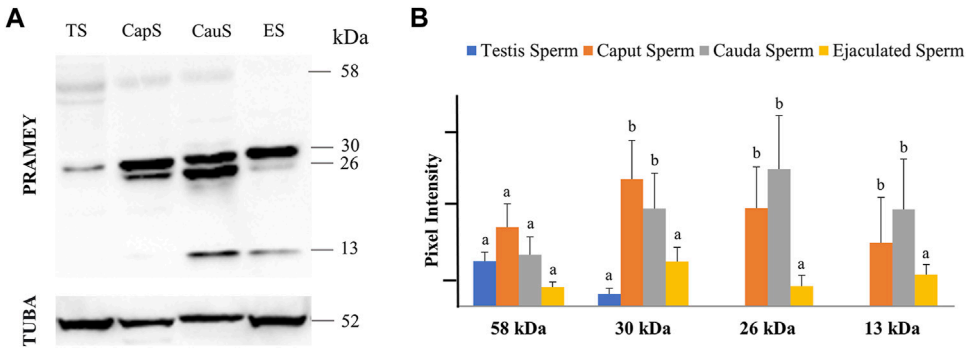


FIGURE 1 | Expression of PRAMEY isoforms in bovine spermatozoa collected from testis (TS), caput (CapS), and cauda (CauS) segments of the epididymis and ejaculated spermatozoa (ES). **(A)** Immunoblot of total sperm protein in each group probed by PRAMEY antibody and then reprobed by α-tubulin (TUBA) antibody. **(B)** Quantification of pixel intensities for the 58, 30, 26, and 13 kDa PRAMEY bands, normalized to the corresponding TUBA. Data are expressed as the mean of the protein pixel intensity ±SEM (n = 5). Significance was evaluated between the four types of sperm (testis, caput, cauda, ejaculated) for each PRAMEY isoform (58, 30, 26, and 13 kDa). Values without a common superscript differed ($p < 0.05$).

TABLE 1 | Summary of PRAMEY isoforms detected in bovine sperm, fluid, and tissue by WB[a].

PRAMEY	58 kDa	30 kDa	26 kDa	13 kDa
Testis sperm	+	+	–	–
Caput sperm	+	+ + + +	+ + +	+ +
Cauda sperm	+	+ + +	+ + + +	+ + +
Ejaculated sperm	+	+ +	+	+
Testis fluid	+	+ +	–	–
Caput fluid	+	+ + +	-	+
Cauda fluid	+	+ +	+	+ +
Seminal Plasma	+	+	–	+ +
Testis tissue	+	+ + +	–	–
Caput tissue	+	+ + +	–	+
Cauda tissue	+	+ + +	–	+ +

[a]*Note: +/–: PRAMEY, isoforms were either detected (+) or not detected (–) by the PRAMEY, antibody; +: low detection; ++: medium detection; +++/++++: high detection.*

antibody (Liu et al., 2017). In the current study, all four isoforms of PRAMEY were found to be present in testicular, epididymal, and ejaculated spermatozoa of mature adult bulls (**Figure 1A**). The 58 kDa protein, which is the predicted molecular weight for the PRAMEY was detected, but at a low amount in spermatozoa (**Table 1**) from testis, caput, and cauda of the epididymis and ejaculated sperm ($p > 0.05$) (**Figure 1A,B**). The 30 kDa isoform had a higher expression in epididymal spermatozoa (**Table 1**) compared to testicular and ejaculated sperm ($p < 0.05$) (**Figure 1A,B**). As spermatozoa moved to the epididymis, two additional isoforms, 26 and 13 kDa were detected. Although there was no difference ($p > 0.05$) in the 26 kDa protein amount between caput and cauda sperm, a significant difference was found between epididymal and ejaculated sperm (P<0.05). The protein level of the 13 kDa isoform was similar ($p > 0.05$) between cauda and caput sperm. Comparable to the 26 kDa isoform, the 13 kDa protein amount was higher in epididymal sperm compared to ejaculated sperm (**Figure 1A,B**). These results suggest that the two smaller isoforms of PRAMEY may be involved in sperm maturation and sperm fertilizing ability. However, it is unknown whether the 26 and 13 kDa proteins are new isoforms or they are products of the 58 and 30 kDa isoform cleavage/degradation.

Distribution of PRAMEY in Testicular and Epididymal Fluid and Seminal Plasma

Evaluation of the fluid collected from the testis, caput, and cauda segment of the epididymis and ejaculate (seminal plasma) revealed a different PRAMEY pattern from that observed for spermatozoa. As shown in **Figure 2A**, a weak expression of the 58 kDa isoform was observed in all fluid types (**Table 1**). Where the 58 kDa expression was similar among the fluids from testis and epididymis, its expression level was lower ($p < 0.05$) in seminal plasma (**Figure 2A,B**). The 30 kDa isoform was strongly detected compared to the 58 kDa and its expression was higher in fluid from testis and caput compared to seminal plasma ($p < 0.05$). Interestingly, the 26 kDa isoform was not observed in fluid from the testis, while it was hardly observed in caput epididymal fluid or seminal plasma, but was mainly observed in fluid from the cauda segment of the epididymis ($p < 0.05$). The 13 kDa isoform was weakly detected in fluid from the caput segment of the epididymis but had a higher ($p < 0.05$) protein concentration observed in fluid from the cauda epididymis and seminal plasma (**Figure 2A,B**) (**Table 1**).

Expression of PRAMEY in the Testicular and Epididymal Tissues

The presence of PRAMEY in the fluids suggested that PRAMEY may have originated from the testicular and/or epididymal tissue. Therefore, we next evaluated the PRAMEY protein isoform expression in testicular and epididymal tissue (**Figure 3A,B**). This evaluation showed that the 58 kDa isoform was relatively weakly expressed in testicular and epididymal tissue. However, its expression level was higher in tissue from the cauda segment of the epididymis compared to tissue from the testis and the caput segment of the epididymis ($p < 0.05$) (**Figure 3A,B**). In contrast

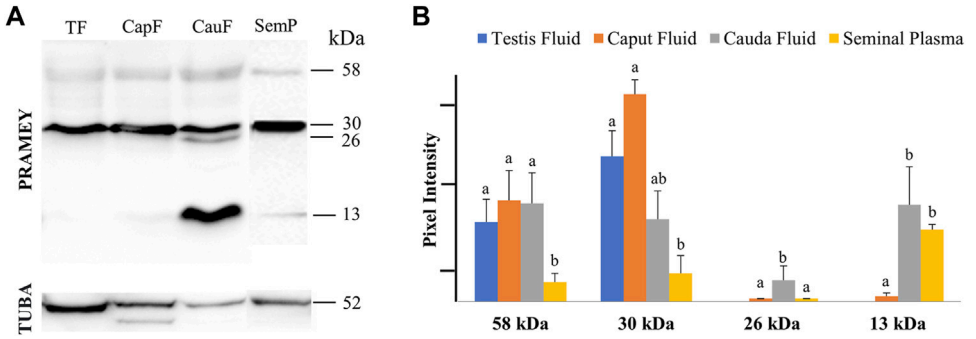


FIGURE 2 | Expression of PRAMEY isoforms in bovine testis (TF), caput epididymal fluid (CapF), cauda epididymal fluid (CauF), and seminal plasma (SemP). **(A)** Immunoblot of total fluid protein in each group probed by PRAMEY antibody and then reprobed by α-tubulin (TUBA) antibody. **(B)** Quantification of pixel intensities for the 58, 30, 26, and 13 kDa PRAMEY bands, normalized to the corresponding TUBA. Data are expressed as the mean of the protein pixel intensity ±SEM for testis ($n = 5$), epididymal fluids ($n = 5$), and seminal plasma ($n = 9$). Significance was evaluated between the four types of fluid (testis, caput, cauda, and seminal plasma) for each PRAMEY isoform (58, 30, 26, and 13 kDa). Values without a common superscript differed ($p < 0.05$).

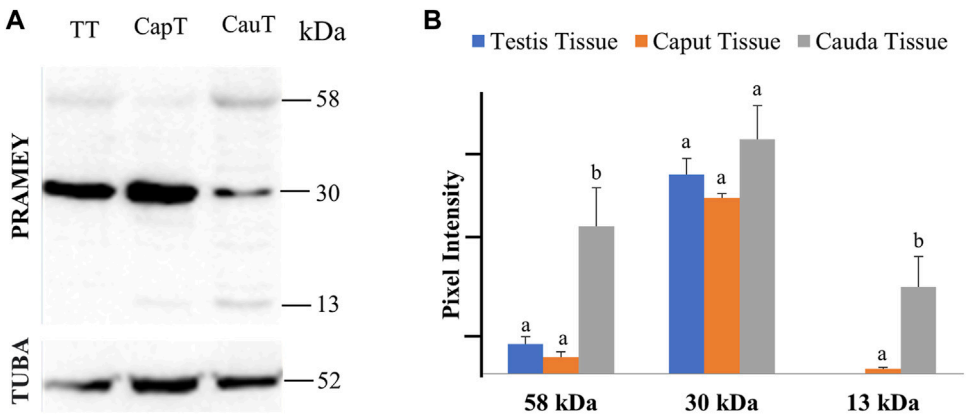


FIGURE 3 | Expression of PRAMEY isoforms in bovine testis tissue (TT) and caput (CapT) and cauda (CauT) epididymal tissues. **(A)** Immunoblot of total tissue protein in each group probed by PRAMEY antibody and then reprobed by α-tubulin (TUBA) antibody. **(B)** Quantification of pixel intensities for the 58, 30, and 13 kDa PRAMEY bands, normalized to the corresponding TUBA. Data are expressed as the mean of the protein pixel intensity ±SEM ($n = 5$). Significance was evaluated between the three types of tissue (testis, caput, cauda) for each PRAMEY isoform (58, 30, and 13 kDa). Values without a common superscript differed ($p < 0.05$).

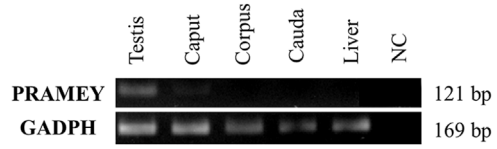


FIGURE 4 | Expression analysis of the bovine *PRAMEY* gene by RT-PCR in bovine testis and epididymal tissues (caput, corpus, and cauda). The liver RNA sample and water (no RNA) were used as negative controls (NC), while the expression of the *GADPH* gene was used as a positive control. PRAMEY was expressed specifically in the testis, while no expression was observed in epididymal and liver tissues.

to the 58 kDa isoform, the 30 kDa isoform was heavily expressed in all three tissues studied (**Table 1**), however, no significant difference was observed among them ($p > 0.05$) (**Figure 3A,B**). Interestingly, the 26 kDa isoform was not observed in the testis or epididymal tissue (**Figure 3A,B**). The 13 kDa protein expression was weakly detected in caput tissue, while its expression level was higher in cauda tissue ($p < 0.05$) (**Figure 3A,B**). Moreover, the 13 kDa isoform was not observed in testicular tissue (**Figure 3A,B**) (**Table 1**).

mRNA Expression of PRAMEY in the Testicular and Epididymal Tissues

To address the question of whether the *PRAMEY* gene is expressed in the epididymal tissue, RT-PCR was performed with a pair of *PRAMEY* gene-specific primers on testicular and epididymal tissues (**Figure 4**). We found that PRAMEY was expressed only in the testis, while no expression was detected in the caput, corpus, or cauda segments of the epididymis, and the negative control, i.e. the liver tissue (**Figure 4**), confirming our

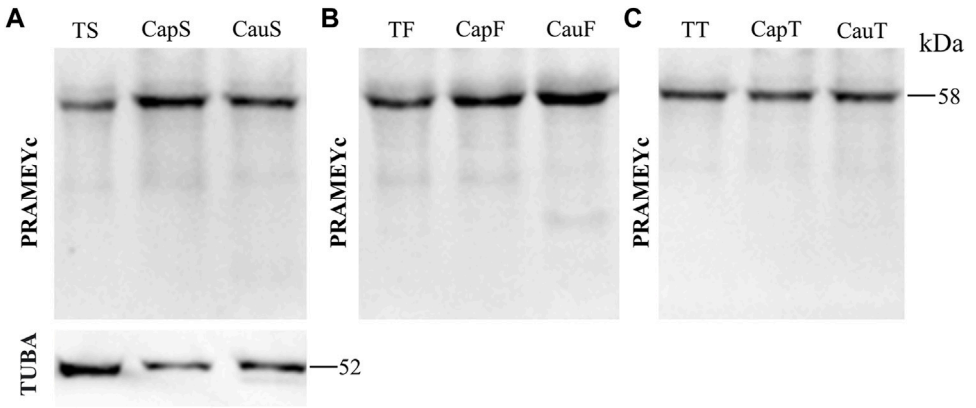


FIGURE 5 | Expression of PRAMEYc in sperm, fluid, and tissue collected from the bovine testis and caput and cauda segments of the epididymis. **(A)** Immunoblot of total testis sperm (TS), caput sperm (CapS), and cauda sperm (CauS) protein probed by the PRAMEYc antibody, then reprobed by α-tubulin (TUBA) antibody. **(B)** Immunoblot of total testis fluid (TF), caput fluid (CapF), and cauda fluid (CauF) protein probed for PRAMEYc content. **(C)** Immunoblot of total testis tissue (TT), caput tissue (CapT), and cauda tissue (CauT) protein probed for PRAMEYc content.

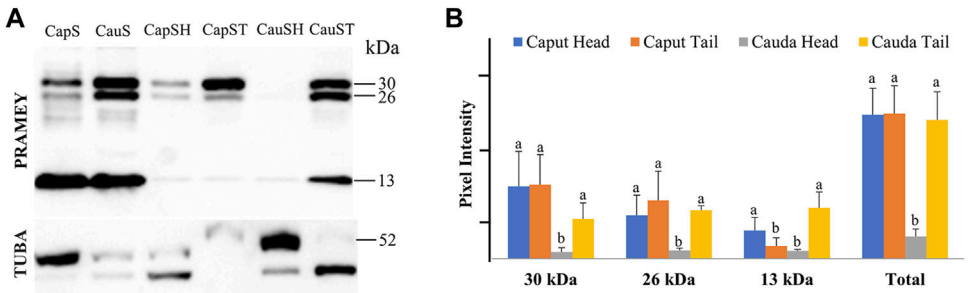


FIGURE 6 | Expression of PRAMEY isoforms in bovine whole caput sperm (CapS), whole cauda sperm (CauS), caput sperm head (CapSH), caput sperm tail (CapST), cauda sperm head (CauSH), and cauda sperm tail (CauST) collected from caput and cauda segments of the epididymis. **(A)** Immunoblot of total sperm protein in each group probed for PRAMEY content and then reprobed for α-tubulin (TUBA). **(B)** Quantification of pixel intensities for the 30, 26, and 13 kDa PRAMEY bands, normalized to the corresponding TUBA. Data are expressed as the mean of the protein density intensity ±SEM ($n = 3$). Significance was evaluated between the four types of sperm (caput head, caput tail, cauda head, and cauda tail) for each PRAMEY isoform (30, 26, and 13 kDa). A different letter indicates a significant difference ($p < 0.05$).

previous report that PRAMEY is germ cell-specific (Liu et al., 2017).

PRAMEYc Expression Pattern in Spermatozoa, Fluids, and Tissue From Testis and Epididymis

The PRAMEY antibody used in the above experiments was referred to as a N-terminus antibody, which produced four isoforms in the WB results. To determine whether a different antibody designed from the C-terminus of PRAMEY would detect the same isoforms, a custom-made C-terminal PRAMEY antibody (called PRAMEYc) was produced and used in this study. We found that the new PRAMEYc antibody detected the intact 58 kDa PRAMEY only, and the expression level of 58 kDa isoforms was low, but consistent, and no significant variations were found among spermatozoa (**Figure 5A**), fluid (**Figure 5B**), and tissue (**Figure 5C**) from testis and caput and cauda segments of the epididymis ($p > 0.05$). The finding indicated that the 30, 26, and 13 kDa PRAMEY isoforms were generated from the N-terminus of PRAMEY.

PRAMEY Expression Pattern in Sperm Head and Tail

We next evaluated the dynamic of PRAMEY expression between sperm head and tail during maturation in the epididymis (**Figure 6A**). Among the four PRAMEY isoforms, the 58 kDa isoform was very weakly detected in the separated sperm head and tail, thus it was not evaluated in this experiment. When comparing the 30, 26, and 13 kDa isoforms individually (**Figures 6A,B**), we found a remarkable decrease of 10.9, 5.4, and 3.8-fold respectively, in protein expression from the caput to cauda epididymis in sperm heads, but there was a small decrease in the expression for

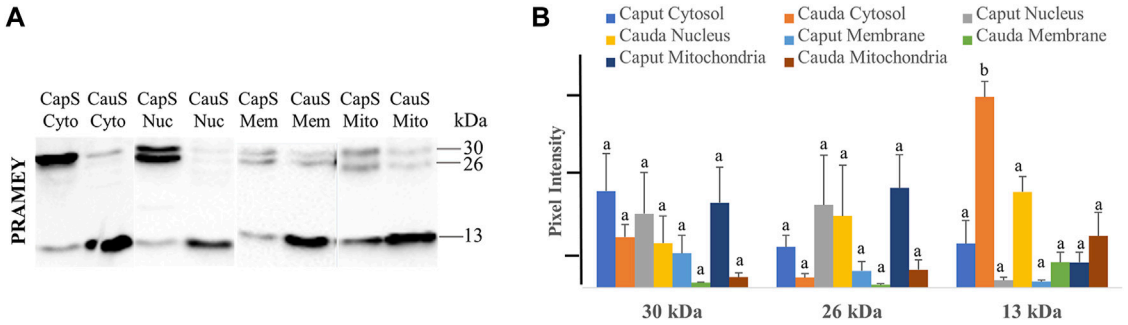


FIGURE 7 | Expression of PRAMEY isoforms in the following subcellular compartments: caput sperm cytosol (CapS Cyto), cauda sperm cytosol (CauS Cyto), caput sperm nucleus (CapS Nuc), cauda sperm nucleus (CauS Nuc), caput sperm membrane (CapS Mem), cauda sperm membrane (CauS Mem), caput sperm mitochondria (CapS Mito), and cauda sperm mitochondria (CauS Mito). **(A)** Immunoblot of total subcellular compartment protein in each group probed by PRAMEY antibody. **(B)** Quantification of pixel intensities for the 30, 26, and 13 kDa PRAMEY bands. Data are expressed as the mean of the protein pixel intensity ±SEM (n = 3). Significance was evaluated between the eight types of subcellular compartments (cytosol, nucleus, membrane, and mitochondria from caput and cauda sperm) for each PRAMEY isoform (30, 26, and 13 kDa). Values without a common superscript differed ($p < 0.05$).

both the 30 and 26 kDa isoforms in sperm tails (1.9 and 1.2-fold respectively). In contrast, the 13 kDa isoform increased 4-fold in sperm tails from caput to cauda (**Figures 6A,B**), suggesting this isoform may have a significant role in tail function, but is likely not as important in sperm head function of cauda spermatozoa. Variations in protein expression were observed among bulls, particularly in the sperm head. For example, the 30 and 26 kDa isoforms in the sperm head of this animal shown in **Figure 6A** were not detected. When expression levels of all isoforms were combined, PRAMEY expression was nearly equivalent for both caput sperm head and tail (**Figure 6B**). Alternatively, cauda sperm tails had an expression 6-fold higher than cauda sperm heads, suggesting the role of PRAMEY in tail function (**Figure 6B**). It is worth noting that α-tubulin antibodies detected two bands from the sperm head and tail samples (**Figure 6A**). This phenomenon also occurred in α-tubulin detected in caput fluid (**Figure 2A**).

Expression of PRAMEY in Subcellular Compartments of Spermatozoa

To better understand the PRAMEY functional dynamic between head and tail during sperm maturation, we further evaluated the PRAMEY expression in sperm cytosol, nucleus, membrane, and mitochondria during spermatozoal maturation in the epididymis (**Figure 7A**). Similar to the head and tail experiment, the 58 kDa isoform was very weakly detected, thus it was not evaluated in this experiment. During bovine sperm maturation in the epididymis, the 30 and 26 kDa isoforms tended to decrease in sperm cytosol, nucleus, plasma membrane, and mitochondria from caput to cauda. However, the 13 kDa isoform significantly increased from caput to cauda sperm cytosol ($p < 0.05$), while the nucleus, membrane, and mitochondria expression tended to increase from caput to cauda spermatozoa (**Figure 7A,B**). The dynamic of PRAMEY localization during sperm transit through epididymis possibly suggests the involvement of PRAMEY in sperm maturation, especially the 13 kDa isoform.

DISCUSSION

Similar to many other CTA genes/families, the function of the *PRAME* gene family is difficult to dissect due to it being an expanded gene family in the eutherian mammals (Chang et al., 2011). One of the biggest challenges when working with a multi-copy gene/protein family is to design a gene/protein-specific antibody. To tackle this challenge, we have made two PRAMEY-specific antibodies using a PRAMEY-specific short peptide either from the N-terminus (referred to as anti-PRAMEY antibody) (Liu et al., 2017) or from the C-terminus (referred to as anti-PRAMEYc antibody), which were used to examine the PRAMEY isoforms during sperm maturation in this study. Among the four PRAMEY isoforms identified by WB, 58, and 30 kDa were detected in testicular spermatozoa, while the 30, 26, and 13 kDa were found in epididymal spermatozoa. This finding raised a question on whether the different isoforms of PRAMEY were encoded by a different variant of PRAMEY paralogs as the copy number of PRAMEY varies on the Y chromosome from 2 to 30 among healthy bulls (Yue et al., 2013) or were the different isoforms cleavage products from the intact (58 kDa) and 30 kDa protein. One of the best ways to address this question is to sequence the PRAMEY isoforms using an immunoprecipitation-mass spectrometry (IP-MS) approach. Unfortunately, we failed to sequence the PRAMEY proteins after several trials of IP-MS analysis. We noticed that other proteomic studies, particularly the one performed on the isolated mouse chromatoid body (CB), could not identify the PRAME-related proteins either (Meikar et al., 2014). Like bovine PRAMEY, mouse PRAMEX1 (X-linked; previously known as PRAME) and PRAMEL1 (on Chr 4) are highly enriched in the mouse CB (Liu et al., 2021). Failure to identify the bovine and mouse PRAME-related proteins in testis/sperm samples by MS may reveal a challenge in sequencing proteins in the PRAME family.

Although the bovine Y chromosome sequence assembly and gene annotation have not been completely finished yet (Chang et al., 2013), the likelihood of post-testicular translation giving

rise to the 26 and 13 kDa from transcript variants of PRAMEY or even from different paralogs of PRAMEY is very low for two possible reasons. One is that there is no evidence to date that the bovine Y chromosome has a copy of *PRAMEY* that encodes a 26 or 13 kDa protein. The other reason is that chromatin of germ cells is undergoing condensation, transcriptional activity is shut down during spermiogenesis, and most of the cytoplasm is lost during sperm maturation (Cooper, 2011; Li et al., 2021; Martins et al., 2021). Therefore, we hypothesize that the 26 and 13 kDa PRAMEY isoforms were a result of cleavage and were released from the spermatozoa into the fluid at some point during spermatozoa maturation.

It was unexpected to find that the 30 kDa PRAMEY protein was heavily present in the fluid of the testis and epididymis, and seminal plasma. The presence of PRAMEY in these fluid samples may be explained by the enrichment of PRAMEY in late the CB and the mitochondria-associated granule (MAG) within the cytoplasm of elongated spermatids (Liu et al., 2017). When the cytoplasm of elongated spermatids is shed at the end of spermiogenesis, the CB, and MAG, which contain PRAMEY, will also be lost from the spermatids through the cytoplasmic droplets. It is believed that the cytoplasmic droplets are released mainly in the caput segment of the epididymis during sperm transition from the testis to cauda epididymis in sperm maturation (Branton and Salisbury, 1947; Cooper and Yeung, 2003; Zhou et al., 2018; Sellem et al., 2021). Therefore, PRAMEY may be shed into epididymal lumen fluid during maturational passage of spermatozoa. Accordingly, the content of 26 and 13 kDa isoforms increases in both epididymal spermatozoa and fluid from caput to cauda. These findings may suggest a transcytosis pathway, where PRAMEY is being transported transcellularly after being released from the spermatid/spermatozoa cytoplasm. If this hypothesis is correct, the PRAMEY protein observed in epididymal fluid and tissue likely originated from spermatozoa (*i.e.* germ cells) during sperm maturation, not from non-germ cells, such as the epididymis. Our RT-PCR results with PRAMEY-specific primers indeed confirmed that the *PRAMEY* gene is expressed in the testis, but not in the epididymal and liver tissues.

Our data from the subcellular compartments of spermatozoa revealed that the 13 kDa PRAMEY isoform increased 4-fold in sperm tails from caput to cauda, suggesting that this isoform may have a significant role in tail function. It is further supported by the data showing that the 13 kDa PRAMEY also increased within mitochondria from caput to cauda sperm, strongly indicating this isoform's role in spermatozoal motility. This is in line with the previous discovery that PRAMEY interacts with PP1γ2 (Liu et al., 2017), a testis/spermatozoa-specific phosphatase that regulates spermatozoal motility (Smith et al., 1996; Huang et al., 2002; Fardilha et al., 2011; Silva et al., 2021).

In conclusion, the 58 and 30 kDa PRAMEY isoforms are the primary isoforms detected in testicular spermatozoa, fluid, and tissue, suggesting their involvement in spermatogenesis (**Table 1**). The heavily expressed 30 and 13 kDa PRAMEY isoforms in epididymal and ejaculated spermatozoa strongly suggest their involvement in sperm maturation and fertilizing capability (**Table 1**). Data from this study and our previous work (Liu et al., 2017; Kern, 2020) supports the proposal that PRAMEY's function is dynamic during spermatogenesis (Liu et al., 2017), sperm maturation (this work), and sperm-acrosomal function and fertilization (Kern, 2020).

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committees of the Pennsylvania State University.

AUTHOR CONTRIBUTIONS

W-SL conceived and designed the research project. CK and WF participated in the experimental design. WF carried out WB analysis on spermatozoa, fluid, and tissue of the testis, epididymis, and fresh semen, as well as sperm subcellular components. CK carried out WB analysis on sperm head and tail and RT-PCR. CK took the lead in writing the manuscript. WL provided critical feedback during manuscript writing and revision. All authors approved the manuscript.

FUNDING

This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2018-67015-27576 from the USDA National Institute of Food and Agriculture (NIFA) to W-SL.

ACKNOWLEDGMENTS

We would like to thank Mr. Douglas Nicholas and Mr. Duane Eichenlaub at Nicholas Meat, LLC (Loganton, PA) for providing bovine testes, and Dr. Lester Griel at the Department of Veterinary and Biomedical Sciences, Penn State University for his help in collecting bull semen. We are also grateful to the Pennsylvania Department of Agriculture Livestock Evaluation Center for providing fresh bovine semen samples.

REFERENCES

Birtle, Z., Goodstadt, L., and Ponting, C. (2005). Duplication and Positive Selection Among Hominin-specific PRAME Genes. *BMC Genomics* 6, 1–19. doi:10.1186/1471-2164-6-120

Branton, C., and Salisbury, G. W. (1947). Morphology of Spermatozoa from Different Levels of the Reproductive Tract of the Bull1. *J. Anim. Sci.* 6, 154–160. doi:10.2527/jas1947.62154x

Chang, T.-C., Yang, Y., Retzel, E. F., and Liu, W.-S. (2013). Male-specific Region of the Bovine Y Chromosome Is Gene Rich with a High Transcriptomic Activity in Testis Development. *Proc. Natl. Acad. Sci.* 110, 12373–12378. doi:10.1073/pnas.1221104110

Chang, T.-C., Yang, Y., Yasue, H., Bharti, A. K., Retzel, E. F., and Liu, W.-S. (2011). The Expansion of the PRAME Gene Family in Eutheria. *PLoS ONE* 6 (2), e16867. doi:10.1371/journal.pone.0016867

Church, D. M., Goodstadt, L., Hillier, L. W., Zody, M. C., Goldstein, S., She, X., et al. (2009). Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. *Plos Biol.* 7 (5), e1000112. doi:10.1371/journal.pbio.1000112

Cooper, T. G., and Yeung, C. H. (2003). Acquisition of Volume Regulatory Response of Sperm upon Maturation in the Epididymis and the Role of the Cytoplasmic Droplet. *Microsc. Res. Tech.* 61 (28–38), 28–38. doi:10.1002/jemt.10314

Cooper, T. G. (2011). The Epididymis, Cytoplasmic Droplets and Male Fertility. *Asian J. Androl.* 13 (1), 130–138. doi:10.1038/aja.2010.97

Costessi, A., Mahrour, N., Tijchon, E., Stunnenberg, R., Stoel, M. A., Jansen, P. W., et al. (2011). The Tumour Antigen PRAME Is a Subunit of a Cul2 Ubiquitin Ligase and Associates with Active NFY Promoters. *EMBO* 30 (18), 3786–3798. doi:10.1038/emboj.2011.262

Epping, M. T., Hart, A. A. M., Glas, A. M., Krijgsman, O., and Bernards, R. (2008). PRAME Expression and Clinical Outcome of Breast Cancer. *Br. J. Cancer* 99 (3), 398–403. doi:10.1038/sj.bjc.6604494

Epping, M. T., Wang, L., Edel, M. J., Carlée, L., Hernandez, M., and Bernards, R. (2005). The Human Tumor Antigen PRAME Is a Dominant Repressor of Retinoic Acid Receptor Signaling. *Cell* 122 (6), 835–847. doi:10.1016/j.cell.2005.07.003

Fardilha, M., Esteves, S. L. C., Korrodi-Gregorio, L., Pelech, S., da Cruz e Silva, O. A. B., and da Cruz e Silva, E. (2011). Protein Phosphatase 1 Complexes Modulate Sperm Motility and Present Novel Targets for Male Infertility. *Mol. Hum. Reprod.* 17, 466–477. doi:10.1093/molehr/gar004

Hess, R. A., and de Franca, L. R. (2009). "Spermatogenesis and Cycle of the Seminiferous Epithelium," in *Molecular Mechanisms in Spermatogenesis*. Editor C. Y. Cheng (New York, NY: Springer New York), 1–15. doi:10.1007/978-0-387-09597-4_1

Huang, Z., Khatra, B., Bollen, M., Carr, D. W., and Vijayaraghavan, S. (2002). Sperm PP1γ2 Is Regulated by a Homologue of the Yeast Protein Phosphatase Binding Protein Sds221. *Biol. Reprod.* 67, 1936–1942. doi:10.1095/biolreprod.102.004093

Ikeda, H., Lethé, B., Lehmann, F., Van Baren, N., Baurain, J.-F., De Smet, C., et al. (1997). Characterization of an Antigen that Is Recognized on a Melanoma Showing Partial HLA Loss by CTL Expressing an NK Inhibitory Receptor. *Immunity* 6, 199–208. doi:10.1016/s1074-7613(00)80426-4

Kern, C. H. (2020). Characterization of PRAMEY Isoforms and Their Involvement in Bovine Sperm Capacitation and Acrosome Reaction. Master's thesis. The Pennsylvania State University.

Kern, C. H., Yang, M., and Liu, W.-S. (2021). The PRAME Family of Cancer Testis Antigens Is Essential for Germline Development and Gametogenesis. *Biol. Reprod.* 105 (2), 290–304. doi:10.1093/biolre/ioab074

Kobe, B., and Kajava, A. (2001). The Leucine-Rich Repeat as a Protein Recognition Motif. *Curr. Opin. Struct. Biol.* 11 (6), 725–732. doi:10.1016/s0959-440x(01)00266-4

Li, Y., Mi, P., Chen, X., Wu, J., Qin, W., Shen, Y., et al. (2021). Dynamic Profiles and Transcriptional Preferences of Histone Modifications during Spermiogenesis. *Endocrinology* 162 (1). doi:10.1210/endocr/bqaa210

Liu, W.-S., Lu, C., and Mistry, B. V. (2021). Subcellular Localization of the Mouse PRAMEL1 and PRAMEX1 Reveals Multifaceted Roles in the Nucleus and Cytoplasm of Germ Cells during Spermatogenesis. *Cell Biosci* 11, 102. doi:10.1186/s13578-021-00612-6

Liu, W.-S., Zhao, Y., Lu, C., Ning, G., Ma, Y., Diaz, F., et al. (2017). A Novel Testis-specific Protein, PRAMEY, Is Involved in Spermatogenesis in Cattle. *Reproduction* 153 (6), 847–863. doi:10.1530/rep-17-0013

Martins, M. C., Gonçalves, L. M., Nonato, A., Nassif Travençolo, B. A., Alves, B. G., and Beletti, M. E. (2021). Sperm Head Morphometry and Chromatin Condensation Are in Constant Change at Seminiferous Tubules, Epididymis, and Ductus Deferens in Bulls. *Theriogenology* 161, 200–209. doi:10.1016/j.theriogenology.2020.12.004

Meikar, O., Da Ros, M., Korhonen, H., and Kotaja, N. (2011). Chromatoid Body and Small RNAs in Male Germ Cells. *Reproduction* 142, 195–209. doi:10.1530/rep-11-0057

Meikar, O., Vagin, V. V., Chalmel, F., Sõstar, K., Lardenois, A., Hammell, M., et al. (2014). An Atlas of Chromatoid Body Components. *RNA* 20, 483–495. doi:10.1261/rna.043729.113

Meuwissen, R. L. J., Offenberg, H. H., Dietrich, A. J., Riesewijk, A., van Iersel, M., and Heyting, C. (1992). A Coiled-Coil Related Protein Specific for Synapsed Regions of Meiotic Prophase Chromosomes. *EMBO* 1 (1), 5091–5100. doi:10.1002/j.1460-2075.1992.tb05616.x

Mishra, S., Somanath, P. R., Huang, Z., and Vijayaraghavan, S. (2003). Binding and Inactivation of the Germ Cell-specific Protein Phosphatase PP1gamma2 by Sds22 during Epididymal Sperm Maturation. *Biol. Reprod.* 69 (69), 1572–1579. doi:10.1095/biolreprod.103.018739

Mistry, B. V., Zhao, Y., Chang, T. C., Yasue, H., Chiba, M., Oatley, J., et al. (2013). Differential Expression of PRAMEL1, a Cancer/Testis Antigen, during Spermatogenesis in the Mouse. *PLoS ONE* 8, e60611. doi:10.1371/journal.pone.0060611

Saburi, S., Nadano, D., Akama, T. O., Hirama, K., Yamanouchi, K., Naito, K., et al. (2001). The Trophinin Gene Encodes a Novel Group of MAGE Proteins, Magphinins, and Regulates Cell Proliferation during Gametogenesis in the Mouse. *J. Biol. Chem.* 276, 49378–49389. doi:10.1074/jbc.m108584200

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 Years of Image Analysis', *Nat. Methods*, 9(7), pp. 671–675. doi:10.1038/nmeth.2089

Sellem, E., Marthey, S., Rau, A., Jouneau, L., Bonnet, A., Le Danvic, C., et al. (2021). Dynamics of Cattle Sperm sncRNAs during Maturation, from Testis to Ejaculated Sperm. *Epigenetics & Chromatin* 14, 24. doi:10.1186/s13072-021-00397-5

Silva, J. V., Freitas, M. J., Santiago, J., Jones, S., Guimarães, S., Vijayaraghavan, S., et al. (2021). Disruption of Protein Phosphatase 1 Complexes with the Use of Bioportides as a Novel Approach to Target Sperm Motility. *Fertil. Sterility* 115 (2), 348–362. doi:10.1016/j.fertnstert.2020.08.013

Smith, G. D., Wolf, D. P., Trautman, K. C., da Cruz e Silva, E. F., Greengard, P., and Vijayaraghavan, S. (1996). Primate Sperm Contain Protein Phosphatase 1, a Biochemical Mediator of Motility1. *Biol. Reprod.* 54, 719–727. doi:10.1095/biolreprod54.3.719

Tateno, H., Kimura, Y., and Yanagimachi, R. (2000). Sonication Per Se Is Not as Deleterious to Sperm Chromosomes as Previously Inferred1. *Biol. Reprod.* 63 (1), 341–346. doi:10.1095/biolreprod63.1.341

Tulsiani, D. R., Abou-Haila, A., Loeser, C. R., and Pereira, B. M. (1998). The Biological and Functional Significance of the Sperm Acrosome and Acrosomal Enzymes in Mammalian Fertilization. *Exp. Cel Res* 240 (151–164), 151–164. doi:10.1006/excr.1998.3943

Wadelin, F., Fulton, J., McEwan, P. A., Spriggs, K. A., Emsley, J., and Heery, D. M. (2010). Leucine-rich Repeat Protein PRAME: Expression, Potential Functions and Clinical Implications for Leukemia. *Mol. Cancer* 9, 1–10. doi:10.1186/1476-4598-9-226

Wang, Z., Xu, X., Li, J. L., Palmer, C., Maric, D., and Dean, J. (2019). Sertoli Cell-Only Phenotype and scRNA-Seq Define PRAMEF12 as a Factor Essential for Spermatogenesis in Mice. *Nat. Commun.* 10 (1), 5196. doi:10.1038/s41467-019-13193-3

Yue, X. P., Chang, T. C., DeJarnette, J. M., Marshall, C. E., Lei, C. Z., and Liu, W.-S. (2013). Copy Number Variation of PRAMEY across Breeds and its Association with Male Fertility in Holstein Sires. *J. Dairy Sci.* 96, 8024–8034. doi:10.3168/jds.2013-7037

Zhou, W., De Iuliis, G. N., Dun, M. D., and Nixon, B. (2018). Characteristics of the Epididymal Luminal Environment Responsible for Sperm Maturation and Storage. *Front. Endocrinol. (Lausanne)* 9, 59. doi:10.3389/fendo.2018.00059

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 in Genetics

ORIGINAL RESEARCH
published: 23 March 2022
doi: 10.3389/fgene.2022.833132

Genetic Parameter Estimation and Whole Sequencing Analysis of the Genetic Architecture of Chicken Keel Bending

Zhihao Zhang[1†], Weifang Yang[2†], Tao Zhu[3], Liang Wang[2], Xiaoyu Zhao[4], Guoqiang Zhao[4], Lujiang Qu[3] and Yaxiong Jia[1]**

[1]Institute of Animal Sciences, Chinese Academy of Agricultural Science, Beijing, China, [2]Beijing General Station of Animal Husbandry, Beijing, China, [3]State Key Laboratory of Animal Nutrition, Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, [4]Hebei Dawu Poultry Breeding Co., Ltd., Hebei, China

OPEN ACCESS

Edited by:
Aixin Liang,
Huazhong Agricultural University, China

Reviewed by:
Mohammed Nasr,
Zagazig University, Egypt
Darrin Karcher,
Purdue University, United States

***Correspondence:**
Lujiang Qu
quluj@163.com
Yaxiong Jia
yaxiongjia@163.com

[†]These authors have contributed equally to this work

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

Received: 10 December 2021
Accepted: 24 February 2022
Published: 23 March 2022

Citation:
Zhang Z, Yang W, Zhu T, Wang L, Zhao X, Zhao G, Qu L and Jia Y (2022) Genetic Parameter Estimation and Whole Sequencing Analysis of the Genetic Architecture of Chicken Keel Bending. Front. Genet. 13:833132. doi: 10.3389/fgene.2022.833132

Bone health is particularly important for high-yielding commercial layer chickens. The keel of poultry is an extension of the abdomen side of the sternum along the sagittal plane and is one of the most important bones. In this study, the keel phenotype of White Leghorns laying hen flocks showed significant individual differences. To clarify its genetic mechanism, we first estimated the heritability of keel bend (KB) in White Leghorn, recorded the production performance of the chicken flock, examined the blood biochemical indexes and bone quality in KB and keel normal (KN) chickens, and performed whole-genome pooled sequencing in KB and KN chickens. We then performed selection elimination analysis to determine the genomic regions that may affect the keel phenotypes. The results show that KB is a medium heritability trait. We found that cage height had a significant effect on the KB ($p < 0.01$). At 48 weeks, there were significant differences in the number of eggs, the number of normal eggs, and eggshell strength ($p < 0.05$). The content of parathyroid hormone was lower ($p < 0.01$) and that of calcitonin was higher ($p < 0.01$) in KB chickens than in KN chickens. The differences in bone mineral density, bone strength, and bone cortical thickness of the humerus and femur were extremely significant ($p < 0.01$), with all being lower in KB chickens than in KN chickens. In addition, the bones of KB chickens contained more fat organization. A total of 128 genes were identified in selective sweep regions. We identified 10 important candidate genes: *ACP5, WNT1, NFIX, CNN1, CALR, FKBP11, TRAPPC5, MAP2K7, RELA,* and *ENSGALG00000047166*. Among the significantly enriched Kyoto Encyclopedia of Genes and Genomes pathways found, we identifed two bone-related pathways, one involving "osteoclast differentiation" and the other the "MAPK signaling pathway." These results may help us better understand the molecular mechanism of bone traits in chickens and other birds and provide new insights for the genetic breeding of chickens.

Keywords: chicken, keel bend, genetic parameters, pool-seq, candidate gene

INTRODUCTION

Laying hens provide eggs and meat for human use and are one of the most important poultry in the global breeding industry. With the significant increase in the demand for poultry eggs (Mueller et al., 2018), commercial layer strains have reached very high egg production levels (Toscano et al., 2020). To form an eggshell of the quality expected by consumers, each egg needs approximately 2–3 g of calcium (Roberts, 2004; Kebreab et al., 2009). The calcium required for the synthesis of eggshells mainly comes from feed (approximately 65%) and bone absorption (about 35%; Mueller et al., 1964; Buss and Guyer, 1984). Bone is the main repository of calcium (Li et al., 2017). When the calcium supply in the feed is insufficient, the endogenous calcium in the bones undergoes reabsorption to meet the needs of daily egg production (Nys and Le Roy, 2018). Therefore, bone health is particularly important for laying hens.

The keel of poultry is an extension of the abdomen side of the sternum along the sagittal plane and is one of the most important bones of poultry. It is an important structure responsible for the flight ability of birds and it plays an important role in breathing (Claessens, 2009; Wei et al., 2020). The keel of poultry is usually the first point of contact when a collision occurs, so it is easily damaged (Donaldson et al., 2012), which is known as keel bone damage (KBD). KBD mainly includes two types, keel fractures and keel bends (KB), which are mainly manifested as bending, offset, or creasing of the keel part or whole. It can also be called keel deformation, keel deviation, and keel fracture, which are all symptoms of KBD (Casey-Trott et al., 2015). It is generally believed that the greatest risk of keel fractures comes from collision between laying hens and their cage and the weakening of keel strength that results (Casey-Trott et al., 2015). KB, defined as a keel with an abnormal shape caused by non-fracture, is a less-mentioned type of keel damage (Casey-Trott et al., 2015). For a normal keel, the front of the keel ridge should follow a straight line; deformation causes a deviation from this straight line. Deviations can be in the vertical or horizontal direction, creating an S-shaped appearance, protrusions, recesses, or other shapes. The terms used to describe such deviations include "S-shaped," 'twisted,' and 'curved' (Lay et al., 2011). Curvature may be caused by the destruction of the keel periosteum surface rather than being a direct result of fracture or impact injury (Fleming et al., 2004). In contrast to fractures, the development of KB takes a period of time, which is the result of the body's response to normal load and pressure, that is, the process of bone remodeling (Riber et al., 2018).

KB affects the physical condition of laying hens, resulting in acute or chronic physiological stimulation of sick laying hens (Nasr M. A. F. et al., 2012; 2013). This is not only an animal welfare issue, but also affects the health status and production performance of laying hens (Nasr M. et al., 2012; Gebhardt-Henrich and Frohlich, 2015) which then affects the economic benefits of the laying hen industry. KB is a trait that is easy to identify during the breeding process. Moreover, studies have reported that the bone characteristics of laying hens can be moderately to highly inherited (Bishop et al., 2000). Because of the high heritability of chicken bone traits and the increasing number of bone problems, research on chicken bone traits at the gene level is gradually increasing. In a recent study, 21 candidate genes that may regulate chicken bone growth and development were identified (Li et al., 2021), namely *LRCH1, RB1, FNDC3A, MLNR, CAB39L, FOX O 1, LHFP, TRPC4, POSTN, SMAD9, RBPJ, PPARGC1A, SLIT2, NCAPG, NKX3-2, CPZ, SPOP, NGFR, SOST, ZNF652*, and *HOXB3*. Previous studies reported genes affecting the tibia and femur of laying hens (Guo et al., 2020), including *HTR2A, LPAR6, CAB39L, TRPC4, WNT9A, SPOP, NGFR, GIP,* and *HOXB3*. Furthermore, candidate genes associated with chicken osteoporosis have also been identified (Guo et al., 2017), including *RANKL, ADAMTS*, and *SOST*.

Compared with other bones of chicken, the keel is completely ossified later (Buckner et al., 1949). Therefore, some areas of the keel remain cartilaginous for a long time (Casey-Trott et al., 2017). Hartmann and Tabin (2000) found that *WNT-5a*, *WNT-5b*, and *WNT-4* genes are expressed in the chondrogenic region of chicken limb. Their experiment also shows that an endogenous (Hartmann and Tabin, 2000). The Wnt signal does indeed function to promote chondrogenic differentiation. At the same time, some research findings suggest a functional role for Wnt signaling throughout embryonic cartilage development (Daumer et al., 2004). In addition, MAPK signaling pathway-related genes *RAC2, MAP3K1, PRKCB, FLNB, IL1R1, PTPN7, RPS6KA, MAP3K6, GNA12*, and *HSPA8* play an important role in chicken tibial dyschondroplasia (Jahejo et al., 2019). Therefore, it is important to explore the genetic basis of KB in poultry.

In this study, we estimated the genetic parameters of KB in White Leghorn laying hens and analyzed the relationship between KB and production performance. We divided chickens into two groups: those with a normal keel ("keel normal"; KN) and those with KB. The blood biochemical indices related to bone metabolism and the differences in the humerus, femur, and keel between the two groups were compared. Finally, we performed whole-genome pooled sequencing (Pool-Seq) on line White Leghorn chickens with an obvious KB or KN. By screening the genomic regions that experienced selective sweeps, genes related to the keel were identified, and the underlying molecular genetic mechanisms were explored.

MATERIALS AND METHODS

Animals and Data Collection

The laying hens used in the experiment were commercial White Leghorns laying hen strains from a Chinese laying hen breeding company. The total number of laying hens was 1,600. All hens were raised in three-layer H-shaped cages in the same chicken house with closed and mechanical ventilation. Single cage feeding was used. Each single cage was 40 cm long, 45 cm wide, 45 cm high at the front, and 38 cm high at the back. The feeding density was approximately six animals per square meter according to the calculation of a single layer of a single cage. During the

TABLE 1 | Classification of KB severity in White Leghorn chickens.

Level[a]	Description of KB[b]
1	Normal: the keel is streamlined and there is no dent or deviation
2	Slight: the keel is slightly sunken or deviated (visual inspection means that it is easy to make mistakes, thus, this needs to be confirmed by palpation)
3	Moderate: the keel is obviously sunken or deviated (palpation results are consistent with the visual results)
4	Severe: the keel is severely sunken or deviated (palpation results are more serious than the visual results)

[a]*Level: Numbers 1, 2, 3, and 4 indicate the degree of keel bend.*
[b]*KB, keel bend.*

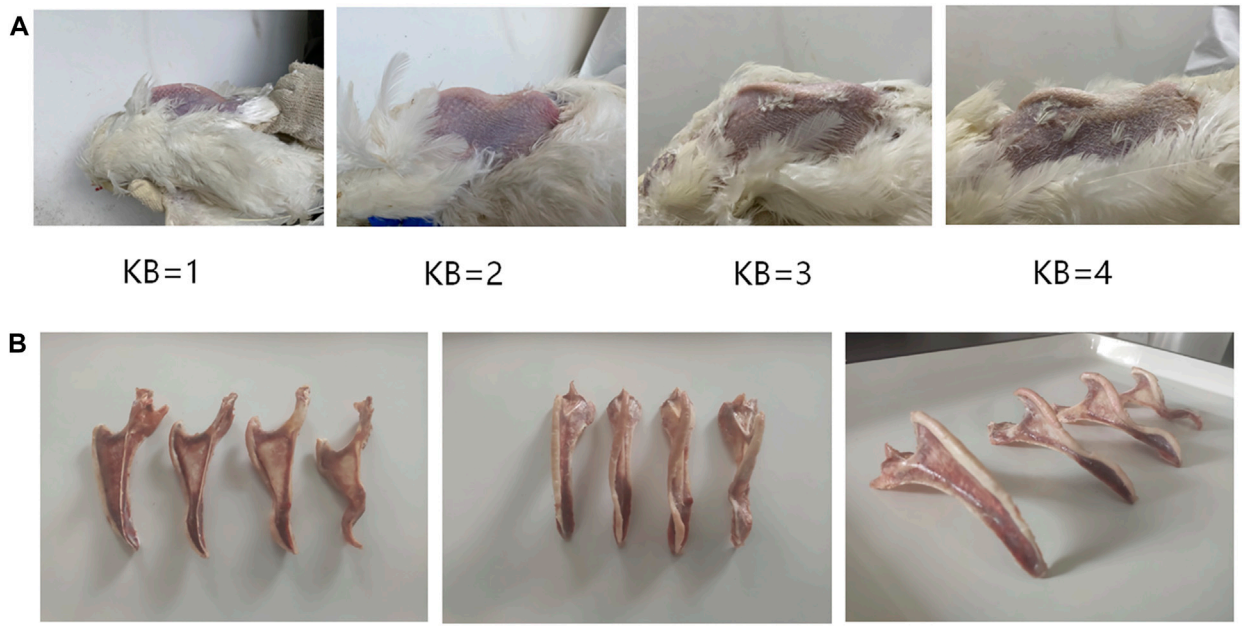


FIGURE 1 | (A) Reference images of KB levels in White Leghorn laying hens **(B)** Anatomical diagram of KB levels in White Leghorn laying hens.

experiment, the same management method and feeding system were used throughout. The KB level of each hen in the experimental flock was determined when the chickens were 30 and 46 weeks old, and the cage height was recorded.

The level of KB was determined using manual palpation. During palpation, the wings of the chicken were held from the wing root in one hand of the palpator to expose the chicken's abdomen. Then, the index finger and thumb of the palpator's other hand were used to examine the ventral and lateral edges of the keel to determine the level of KB. To improve the efficiency of evaluation, we developed an easy and effective method for determining the level of KB. The grading criteria of our method for evaluating KB levels are summarized in **Table 1**. In summary, KBs can be categorized into one of four levels according to the degree of the bend: normal, slight, moderate, or severe, recorded as 1, 2, 3, and 4, respectively (**Figure 1A**). Examples of the different levels are shown in **Figure 1B**. To ensure consistency in assessment, only one palpator was used, and the palpator assessed 100 chickens as a training exercise prior to data gathering. These chickens were assessed many times until the consistency rate of the palpation results of the last two evaluations reached about 90%. In order to study the relationship between KB and production performance in laying hens, we counted the total number of eggs and the total number of "normal" eggs (i.e., excluding those that had a double yolk, soft shell, unsmooth shell, broken shell, or were deformed) of all experimental laying hens from the beginning of laying to 48 weeks (336 days) old. At 48 weeks, we collected the eggs produced by all experimental laying hens on a given day and measured their weight and eggshell strength. To measure the eggshell strength, the egg is placed in the upward position and an eggshell strength tester (HAD-A1, Beijing Hengaode Instrument Co., Ltd., Beijing, China) was used for measurement.

Genetic Parameters Estimation

In this study, the DMU (v6) software package was used to estimate the components of genetic variance (Madsen et al., 2006). The rjmc module was used to analyze the KB data of the 30- and 46-week-old chickens. The calculation method

used was the Bayesian method based on Gibbs sampling (Ødegård et al., 2010). In this study, the heritability of keel traits was calculated using a univariate animal model. The variance component was estimated using the following formula:

$$y = Xb + Za + e$$

where y is the phenotypic value of chicken KB; X and Z are the correlation matrices of the fixed effect and the random additive effect, respectively; b is the fixed effect vector, including cage height; a is the random additive effect vector; and e represents the random residuals.

Blood Biochemical Indices Related to Bone Metabolism

At 50 weeks of age, we randomly selected 10 KN chickens and 10 KB chickens from the experimental chicken flock. We collected 5 ml of blood from the wing vein of each hen using vacuum blood collection vessels. The serum levels of calcium, phosphorus, parathyroid hormone (PTH), calcitonin (CT), and 1,25-dihydroxyvitamin D3 were measured. To reduce error, measurements were taken twice and average values were used. After blood collection, the left and right femur, left and right humerus, and keel of the chickens were dissected and collected for subsequent index measurements.

Bone Mineral Density, Bone Mineral Content, and Bone Strength

The bone mineral density (BMD) and bone mineral content (BMC) of the right femur, right humerus, and keel were measured using a dual-energy X-ray absorptiometry device (XR-36, Norland, United States). After the tests, the samples were stored at −20°C for subsequent testing. After thawing at 15°C for at least 12 h, the right femur and right humerus samples were broken using the animal bone three-point bend test. A Y019 physical tester (TA. XT plus, Stable Micro Systems, United Kingdom) was used to take measurements by placing the plane of the bone sample downward and horizontally on the two supports of the test device. During measurement, the bone sample was gently held in place by hand to avoid it slipping to improve the accuracy of the data. The length and diameter of the bone sample were measured, and after many adjustments, the distance between the brackets was set to 40 mm. The pressure tool applied a constant vertical force and moved vertically to the midpoint of the bone at a speed of 10 mm/min. Finally, the bone was broken and the ultimate force F required to break the bone was recorded in Newtons to obtain the maximum flexural strength.

The data for each index measured in this experiment were analyzed using R (version 4.0.2) software. One-way analysis of variance (ANOVA) and multiple comparisons were used to analyze the differences among hens of the 4 KB levels.The results are expressed as the mean ± SEM. The levels of significance are as follows: $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.

Bone Slice

For the left femur and left humerus, we collected samples of around 2 cm in length from their epiphysis and diaphysis. The bone samples were immediately immersed in 4% paraformaldehyde. After fixation for 24 h, decalcification for 2 weeks, paraffin embedding, sectioning, hematoxylin and eosin staining, and scanning with an automatic digital section scanner (KF-PRO-120, Ningbo Jiangfeng Biological Information Technology Co., Ltd., Ningbo City, China) was performed. The diaphysis part was cross cut and the epiphysis part was longitudinally cut. After the cross section of the diaphysis was scanned with a scanner, the bone density thickness of the humeral diaphysis and the femoral diaphysis was obtained by measuring the scanned image of the diaphysis.

Whole-Genome Sequencing

Laying hens were selected according to KB level. At 52 weeks of age, 36 KN chickens and 48 obvious KB chickens were selected. A 2 ml blood sample was collected from the jugular vein of each individual and stored at −80°C. Genomic DNA was extracted using a genomic DNA extraction kit (DN02, Beijing Aidlab Biotechnologies Co., Ltd., Beijing, China). We then used 1% agarose gel and an ultraviolet spectrophotometer to identify the integrity and purity of the genomic DNA and a TE buffer to adjust the concentration of each DNA sample to 50 ng/uL.

The genomic DNA of the 36 KN chickens was mixed in equal amounts to construct a KN pool. In the same way, the genomic DNA of the 48 KB chickens was mixed in equal amounts to construct a KB pool. Library construction and sequencing were performed at Beijing Biomarker Technologies Co., Ltd. to complete the preparation of the two chicken genomic DNA libraries, and ILLUMINA HISEQ 2500 (Illumina, United States) was used for sequencing. The sequencing depth of each individual was 5x, that is, the sequencing depth of the KN pool was 180x, and the sequencing depth of the KB pool was 240x.

Quality Control and Data Processing

We obtained the original sequence by sequencing contained low-quality reads with adapters. To ensure the quality of the information analysis, raw reads were filtered to obtain clean reads for subsequent information analysis. Only autosomal markers with clear physical location information were used in the analysis. In this study, fastp (v0.19.4) was used to control the quality of the original data. The main steps of data filtering were as follows: 1) reads with adapters were removed; 2) reads with an N content of more than 10% were filtered; and 3) reads containing more than 50% of the bases with a mass value of less than 10 were removed. After obtaining clean data, we used BWA (v0.7.17) to compare them to the chicken reference genome (GRCg6a). The chicken reference genome sequences were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.5/). We used GATK (v3.6) for mutation detection to obtain the vcf file.

Detection of Selective Sweeps

To explore the selection sweep regions related to KB, we adopted a selection elimination analysis based on the population fixation

TABLE 2 | The genetic parameters of KB.

	30-week-old chickens	46-week-old chickens
Add	0.043	0.063
SE^1	0.005	0.009
Residuals	0.120	0.201
SE^2	0.006	0.011
h^2	0.263	0.238

SE^1, stand error of additive effect. SE^2, stand error of residuals effect.

coefficient (F_{ST}). The population fixation coefficient reflects the heterozygosity level of the population alleles and is used to estimate the difference between the average heterozygosity between subpopulations and the average heterozygosity of the entire population. First, we calculated the F_{ST} value for each single nucleotide polymorphism (SNP) site using a custom python script. Then, we used a 50 kb interval as the sliding F_{ST} window, and the 50% overlap interval as the sliding step size. We calculated the sum of the F_{ST} values of all SNP sites in each window and divided it by the number of SNPs in each window to obtain the average F_{ST} of each window value, using the following formula:

$$F_{ST} = (H_T - H_S)/H_T$$

where H_T represents the average heterozygosity of the compound population. H_S represents the average heterozygosity in the subpopulation. We used the F_{ST} mean of the top 1% window as the threshold. Windows larger than this threshold were regarded as candidate regions for selective sweeps.

Candidate Genes and Functional Annotation

Gene position information was annotated using BioMart in the Ensembl database (Ensembl Genes 104). Candidate genes were searched in a window with F_{ST} values >0.1. To determine the possible functions of genes located in the selective scanning region, we converted chicken genes into human homologous genes, uploaded these homologous genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology based annotation system (KOBAS), and used KOBAS-i (Bu et al., 2021) to perform analysis of gene ontology (GO), the KEGG pathway, and KEGG disease. KOBAS is a web server for gene/protein functional annotation (the annotation module) and functional set enrichment (the enrichment module). Given a set of genes or proteins, it can determine whether a pathway, disease, or GO term shows statistical significance. It provides curated sequences and KEGG pathway knowledge for 5,944 species and GO annotations for 71 popular research species. In KOBAS, Fisher's exact test was used as the data statistical method, and the Benjamini and Hochberg method was used as the False Discovery Rate correction method.

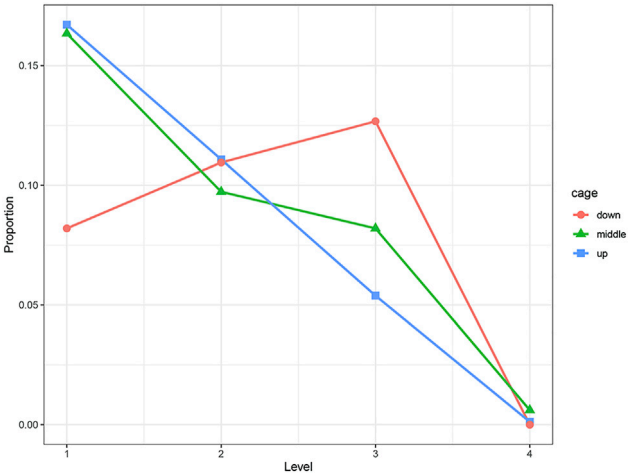


FIGURE 2 | Distribution of KB levels in different cages, where "down" is the first floor, "middle" is the second floor, and "up" is the third floor.

RESULTS

KB Is a Medium Heritability Trait

In this study, we selected approximately 1,600 hens as candidates for phenotypic analysis. According to the DMU instructions, Gibbs sampling method was used to estimate the genetic parameters of KB, and the heritability of 30- and 46-week-old hens was found to be 0.26 and 0.24, respectively (**Table 2**). Our results show that KB is a heritable trait with medium heritability.

KB Correlates With Cage Height and Production Performance

We used the Chi-squared test to examine the effect of cage height on the keel. The results show that cage height has a significant effect on the KB (at 30 weeks of age, $p < 0.01$; at 46 weeks of age, $p < 0.01$). Compared with the middle and upper cages, the proportion of level 1 hens in the bottom cage was significantly less. The proportion of level 2 hens was almost the same, and the proportion of level 3 hens was significantly more. The trends of hens of each level in the middle and upper cages were roughly the same. The proportions of hens in levels 1 to 4 decreased in sequence, and the proportions in each level were similar (**Figure 2**).

The results of the one-way ANOVA showed that there were significant differences in the total number of eggs, total number of normal eggs, and eggshell strength at 48 weeks ($p < 0.05$), but no significant difference in egg weight (**Table 3**). The results of multiple comparisons showed that there was no significant differences in egg weight at 48 weeks among the four levels. The total number of eggs and the total number of normal eggs at 48 weeks in level 4 hens were significantly different from that of the other three levels, with the number of eggs and number of normal eggs both lower (**Table 3**).

TABLE 3 | Production performances of laying hens at different KB levels.

Item[a]	KB[b] = 1	KB = 2	KB = 3	KB = 4	*p*-value[e]
EW	58.96[c] ± 0.13[ad]	58.86 ± 0.16[a]	58.64 ± 0.18[a]	59.50 ± 0.84[a]	0.486
ES	27.68 ± 0.21[ab]	28.09 ± 0.26[b]	27.04 ± 0.29[a]	30.41 ± 2.78[ab]	0.018*
EN	197.36 ± 0.44[b]	197.54 ± 0.50[b]	196.0 ± 0.56[b]	185.00 ± 4.00[a]	<0.001***
NEN	193.63 ± 0.46[b]	193.97 ± 0.52[b]	192.72 ± 0.58[b]	183.00 ± 4.03[a]	0.011*

[a]*EW, egg weight at week 48; ES, eggshell strength at week 48; EN, total number of eggs from the beginning of laying to 48 weeks of age; NEN, total number of normal eggs (i.e., excluding those that had a double yolk, soft shell, unsmooth shell, broken shell, or were deformed) from the beginning of laying to 48 weeks of age.*

[b]*KB, keel bend. 1, 2, 3, and 4 indicate the degree of KB.*

[c]*Data represent the mean ± SEM.*

[d]*Different letters indicate a significant difference, and the same letters indicate no significant difference (p > 0.05).*

[e]**p < 0.05, **p < 0.01, ***p < 0.001, p > 0.05.*

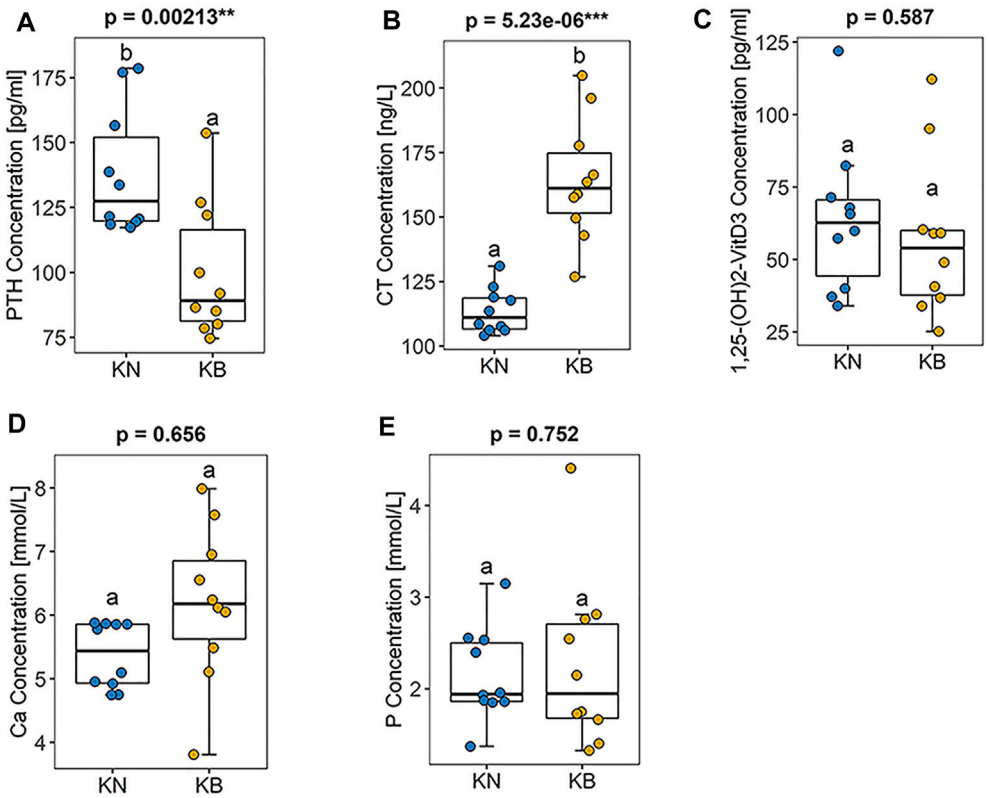


FIGURE 3 | Boxplots of blood biochemical indexes in KN and KB chicken groups **(A)** PTH concentration **(B)** CT concentration **(C)** 1,25-dihydroxyvitamin D3 concentration **(D)** Ca concentration, and **(E)** P concentration.

Blood Biochemical Indices

There were significant differences in the contents of PTH ($p < 0.01$) and CT ($p < 0.01$) between KB and KN chickens. The PTH content was lower and the CT content was higher in KB chickens (**Figures 3A,B**). There were no significant differences in serum calcium ($p > 0.05$), serum phosphorus ($p > 0.05$), and 1,25-dihydroxyvitamin D3 ($p > 0.05$) levels between the two groups (**Figures 3C–E**).

BMD, BMC, Bone Strength, and Bone Cortical Thickness

There were significant differences in humeral BMD, humeral BMC, humeral bone strength, and humeral bone cortical thickness between the two groups ($p < 0.01$). These indexes were higher in KN chickens than in KB chickens (**Figure 4A**). There were significant differences in femoral BMD, femoral humeral bone strength, and femoral bone cortical thickness between the two groups ($p < 0.01$), but there was no significant difference in BMC. For the femur, except BMC, the indexes of KN chicken were higher than those of KB chicken (**Figure 4B**). The above indexes of KB chickens were poor. There was no significant difference in keel BMC or keel BMD between the two groups ($p > 0.05$; **Figure 4C**).

Bone Slice

We sectioned the left humerus and left femur of KB and KN chickens. The diaphysis part was cross cut, and the epiphysis part was

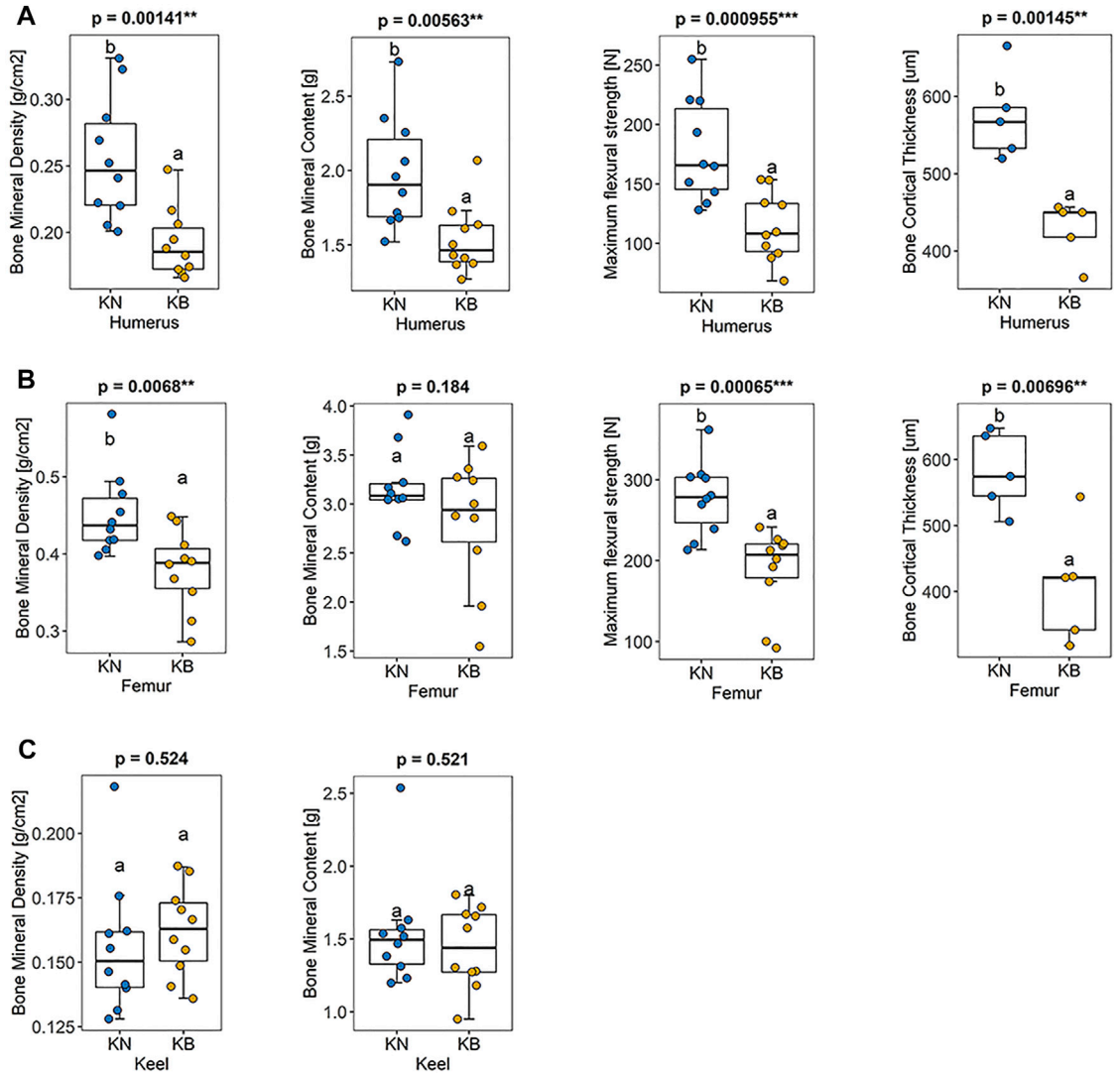


FIGURE 4 | Boxplots of bone mineral density, bone mineral density, bone strength, and bone cortical thickness in KN and KB chicken groups **(A)** Bone index of the humerus **(B)** bone index of the femur, and **(C)** bone index of the keel.

longitudinally cut; the difference between the two groups is shown in **Figure 5**. For the femur, the femoral diaphysis of KN chickens had a bone trabecular structure and a small amount of fat (**Figure 5A**). The femoral diaphysis of KB chickens had almost no trabecular structure or more adipose tissue (**Figure 5B**). The femoral epiphysis of both groups had a trabecular bone structure, but the femoral epiphysis of KB chickens contained more adipose tissue (**Figures 5C,D**). For the humerus, there were no significant differences in humeral diaphysis between the two groups. The humeral epiphysis of the two groups had a trabecular structure, but there was more adipose tissue in the humeral epiphysis of KB chickens (**Figures 5E,F**).

Genome Sequencing Data

After quality control, approximately 1,602.40 Mb of clean reads and 480.00 Gb of clean data were obtained, and the Q30 reached an average of 91.76%. The evaluation results of the sequencing output data for each pool are shown in **Table 4**. The chicken genome was approximately 1.0 GB, with a total of 84 individuals and an average sequencing depth of 5.7x, which meets sequencing requirements. After using the GATK software (v3.6) for mutation detection, the KN group obtained 3.63G data and the KB group obtained 3.76G data.

Selective Sweep Analysis

We screened out 86 regions that may have been affected by selection, as determined by the F_{ST} value (**Supplementary Table S1**). The mean F_{ST} of the top 1% window was 0.097; therefore, we set the threshold to 0.1. One hundred and twenty-eight genes were discovered using BioMart in the Ensembl database. These genes may be related to KB (**Supplementary Table S2**). A Manhattan plot of F_{ST} results is shown in **Figure 6A**.

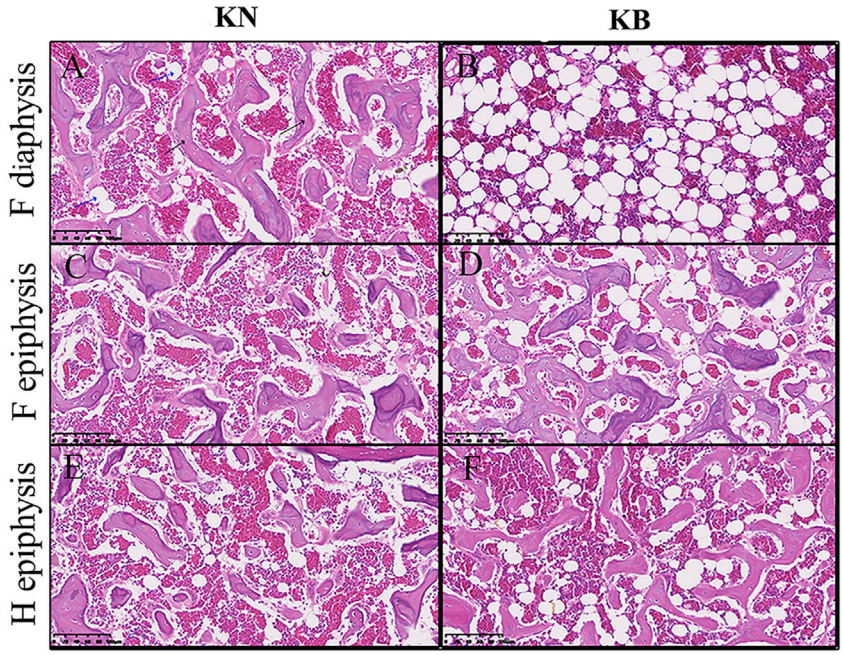


FIGURE 5 | Bone slice chart of the humerus and femur in KN and KB chicken groups. F, femur; H, humerus **(A and B)** Femur diaphysis **(C and D)** femur epiphysis, and **(E and F)** humerus epiphysis. The black arrow indicates the bone trabecular structure. The blue arrow (vacuolated) indicates the adipose tissue.

TABLE 4 | Evaluation statistics of sequencing data of keel normal (KN) and keel bend (KB) chicken groups.

Sample	KN	KB
Raw_Reads	724,883,017	880,771,276
Clean_Reads	723,196,549	879,203,139
Clean_Bases	216,631,412,100	263,366,444,500
Q20 (%)	96.57	96.74
Q30 (%)	91.57	91.95
GC (%)	42.51	42.49

Using Ensembl annotations and based on the function, phenotype description, and previous research reports of these genes, we identified 10 important candidate genes related to keel (**Table 5**). Among them, nine were located on chromosomes 30, 31, and 33 (**Figure 6B**), and one was located on chromosome 1. In addition, the PCSK2 found on chromosome three could also be related to bone quality.

GO and KEGG Pathway Enrichment Analysis

To determine the possible functions of genes located in the selective scanning region, we converted these gene IDs into human homologous genes to obtain 89 human homologous gene IDs. These were then sent to KOBAS for enrichment analysis (**Supplementary Table S2**). GO and KEGG pathway analyses were used to determine the biological functions of these genes. We found 11 significantly enriched KEGG pathways (**Figure 7**). In addition, musculoskeletal diseases and congenital malformations were found in the KEGG disease. Detailed information regarding the analysis of GO, the KEGG pathway, and KEGG disease is shown in **Supplementary Table S3**.

DISCUSSION

Heritability of KB

In this study, we treated KB as a threshold trait by assigning numbers one to four according to the extent of the bend of the keel. A similar strategy has been used in other studies for other factors, such as chicken manure moisture content (Zhu et al., 2020), chicken feather peaking (Brinker et al., 2014), and cow body condition (Edmonson et al., 1989). For the first time, we estimated the heritability of KB. We determined that KB is a heritable trait with medium heritability, indicating that KB is affected by genetic factors. Genetic selection could be an effective method for changing the KB level in the chickens.

We found that the height of the cage was related to the level of KB. The laying hens in the bottom cage showed more severe KB, indicating that the height of the cage is an important factor affecting the level of KB. We speculate that the height of the cage determined many environmental factors, such as air freshness, light intensity, temperature, and humidity. Therefore, raising laying hens in higher cages could be an effective way to reduce KB. At 48 weeks of age, the total number of eggs and the total number of normal eggs of level 4 layers were significantly

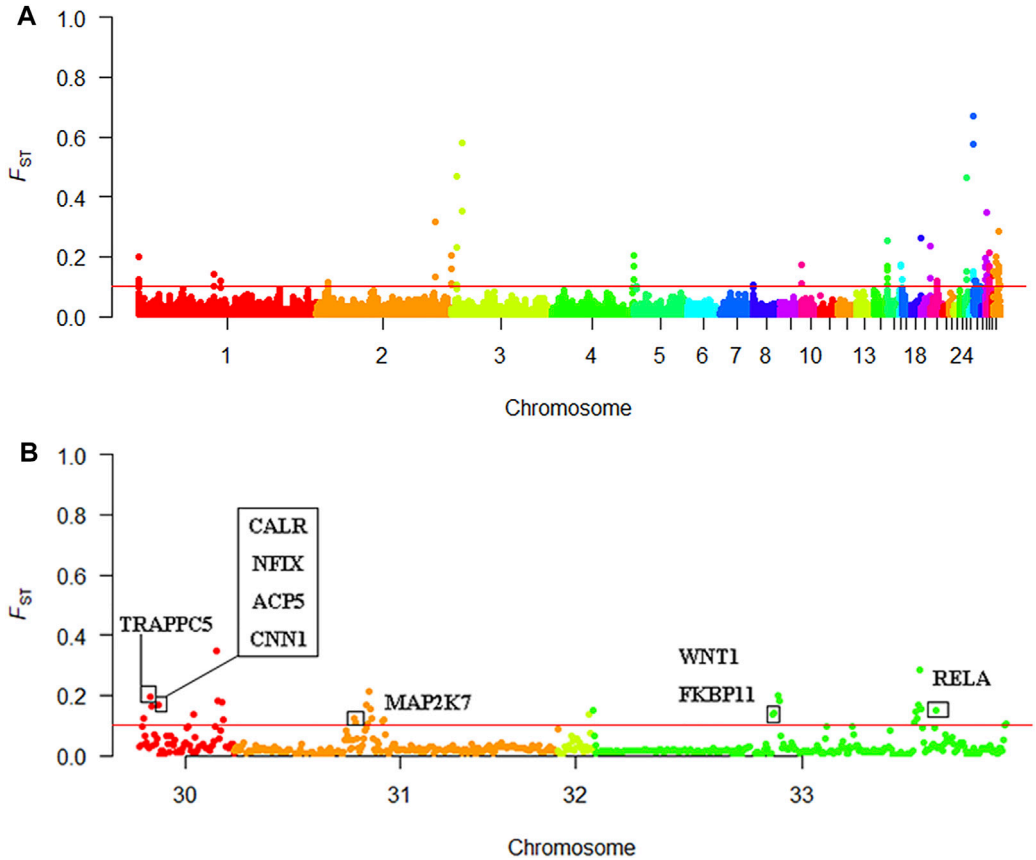


FIGURE 6 | Selective sweep analysis results **(A)** Manhattan plot of the F_{ST} results **(B)** Manhattan plot of chromosomes 30–33.

TABLE 5 | List of important candidate genes associated with keel bend.

Chromosome	Gene start (bp)	Gene end (bp)	Window F_{ST} value	Gene name
1	930,902	935,155	0.20	ENSGALG00000047166
30	422,263	424,991	0.17	CALR
30	402,432	415,709	0.17	NFIX
30	397,120	399,963	0.17	ACP5
30	392,257	395,386	0.17	CNN1
30	241,789	247,265	0.20	TRAPPC5
30	89,765	111,142	0.12	DNM2
31	2,309,015	2,328,588	0.12	MAP2K7
33	3,422,447	3,427,906	0.14	WNT1
33	3,412,284	3,412,858	0.14	FKBP11
33	6,490,052	6,498,311	0.15	RELA

different from the other three levels, and both were lower. Therefore, our research shows that KB has an impact on the later production performance of laying hens, and severe KB will reduce the production performance of laying hens. Because severe KB usually contain fractures that are not easily detected by palpation. Some results show that normal hens lay more eggs than hens with broken keels (Wei et al., 2020), which is consistent with the results of this study. This may be related to the physiological stress caused by fracture, because fracture will reduce the production performance of laying hens (Mumma et al., 2006).

Blood Biochemical Indices

PTH and CT play important roles in maintaining calcium and phosphorus balance and regulating bone metabolism. PTH can promote bone resorption and turnover, mobilize bone calcium

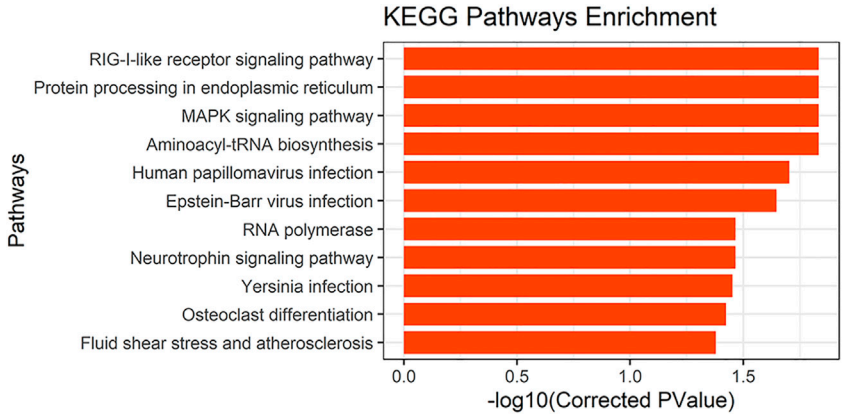


FIGURE 7 | KEGG pathways results.

into the blood, promote the reabsorption of calcium in renal tubules and the absorption of calcium ions in the small intestine, and increase blood calcium and decrease blood phosphorus levels (Goltzman, 2018; Zheng et al., 2020). The main physiological function of CT is to reduce the number of osteoclasts, inhibit the activity of osteoclasts, reduce bone resorption, inhibit the absorption of calcium ions in the small intestine, reduce the concentration of blood calcium *in vivo*, and deposit free calcium in the blood into bone tissue (Naot et al., 2019). There were significant differences in PTH and CT levels between the two groups. The PTH content of KB chickens was lower and the CT content was higher. We speculate that there was relatively little endogenous calcium (calcium in bone) due to the poor bone quality of KB chickens. In order to avoid further loss of endogenous calcium, KB chickens reduce their blood calcium concentration by reducing PTH content and increasing CT content. This maintains a certain level of bone quality while ensuring the amount of calcium required for laying eggs. However, our experimental results showed that there was no significant difference in serum calcium between the two groups, which was different from our prediction. Probably because we are sampling during the day. Because for laying hens, the formation of eggshells mostly occurs at night, and the changes in blood calcium levels may be different at this time. Therefore, the changes of serum calcium levels in laying hens at different times of the day require further study.

Bone Quality

There were significant differences in almost all bone indices for the humerus and femur of the two groups, with the bone indices of KB chickens being lower than those of KN chickens. The results show that the quality of the bones of the KB chickens was poor, and the keel phenotype may be related to bone quality. This is consistent with previous findings that the fracture strength of the humerus and tibia in chickens with normal keels is greater than that in chickens with abnormal keels (Fleming et al., 2004). However, for the keels of the two groups, there was no significant difference in BMD or BMC, except for the obvious difference in phenotype. Some studies have also reported that there is no difference in keel BMD between individuals with and without fractures (Gebhardt-Henrich et al., 2017). The reasons for this need to be studied further. According to our results, there are significant differences in humerus and femur between KN and KB chickens, but no significant differences in the keel. We believe that the use of endogenous calcium in the different bones of laying hens differs in order to maintain high egg yield. We speculate that laying hens use endogenous calcium in the humerus, femur, and other long bones much more frequently than in the keel. This direction is worthy of in-depth study in the future.

In our study, KB chickens had almost no bone trabecular structure in the femur, which may indicate that KB chickens transfer more medullary bone from the femur for eggshell formation. The bones of hens are mainly composed of cortical, cancellous, and medullary bones. In the long bones (femur, humerus, and tibia, for example), a trabecular structure similar to cancellous bone can be formed periodically, that is, the medullary bone (Dominguez-Gasca et al., 2019). The medullary bone is an unstable intimal bone with low collagen fiber content which mainly exists in the cavity of the long bone (Kerschnitzki et al., 2014). The medullary bone provides an unstable source of calcium for eggshell formation, which calcifies and metabolizes much faster than cortical bone (Whitehead, 2004). The medullary bone is also a unique bone tissue formed only in female birds during the breeding period (Rodriguez-Navarro et al., 2018). It is degraded by the transfer of calcium to the eggshell to supplement the insufficient absorption of calcium and phosphorus in the digestive tract (Hudson et al., 1993).

The femur and humerus of KB chickens contained more adipose tissue, and the BMD of the femur and humerus was also lower. An inverse relationship between BMD and bone marrow adipocyte level has been documented in animal and human studies (Martin and Zissimos, 1991; Sabatakos et al., 2000; Wehrli et al., 2000; Shen et al., 2007; Di Iorgi et al., 2008; Shen et al., 2012b). This relationship has been attributed to the ability of mesenchymal stem cells to differentiate into adipocytes or

osteoblasts (Owen and Friedenstein, 1988; Rosen and Bouxsein, 2006). Bone marrow contains different cell populations belonging to several lineages, including hematopoietic stem cells, and mesenchymal stem cells, which can differentiate into osteoblasts, adipocytes, fibroblasts, chondrocytes, and muscle cells (Di Iorgi et al., 2010). Some studies have suggested that the differentiation of bone marrow mesenchymal stem cells into osteoblasts or adipocytes is competitive (Shen et al., 2012a). Similar to other fat pools, bone marrow fat produces adipokines and fatty acids, which might produce a lipotoxic environment in bone cells (Demontiero et al., 2012). Adipocytes inhibit osteoblast proliferation (Maurin et al., 2002) and promote osteoclast differentiation (Hozumi et al., 2009).

Selective Sweep Analysis

Based on the different phenotypes of keel, we carried out whole-genome Pool-Seq of KN and KB chickens. Selective elimination analysis was then performed to identify genes that could lead to different keel phenotypes. This method has been widely used in SNP and functional gene mining and has achieved good results (Kijas, 2014; Rochus et al., 2018). Our analysis was based on the F_{ST}. F_{ST} analysis indicates the degree of population differentiation; the greater the value, the greater the degree of population differentiation and the higher the degree of selection. To narrow the screening range and improve the screening efficiency, we chose F_{ST} = 0.1 as the screening threshold. According to gene function annotation and literature, we identified 10 genes with strong selection signals and bone-related genes, namely nine genes with identified functions and one new gene.

ACP5 encodes tartrate-resistant acid phosphatase (TRACP). TRACP is an abundant protein in osteoclasts, macrophages, and dendritic cells, and its primary substrate is osteopontin (Bilginer et al., 2016). This enzyme is a good marker for bone resorption and osteoclast activity (Bilginer et al., 2016). *ACP5* mutations cause deficient TRACP activity, which results in bone dysplasia through impaired cartilage resorption, particularly at the metaphyses (Lausch et al., 2011). Diseases associated with *ACP5* include spondyloenchondrodysplasia with immune dysregulation (Hayman et al., 1996; Girschick et al., 2015; Briggs et al., 2016; Ramesh et al., 2020). *WNT1* is a member of the Wnt gene family. The Wnt gene family consists of structurally related genes that encode secreted signaling proteins. These proteins have been implicated in oncogenesis and several developmental processes, including regulation of cell fate and patterning during embryogenesis (Zhan et al., 2017). In addition, the Wnt family of proteins drives the development and maintenance of many tissues, including bone (Lu et al., 2018). Numerous human genetic studies and genetically modified mouse models have demonstrated that the Wnt signaling pathway plays an essential role in the regulation of bone formation and resorption (Maeda et al., 2019). *WNT1* plays a role in osteoblast function, bone development, and bone homeostasis (Joeng et al., 2017; Wang et al., 2021). Diseases associated with *WNT1* include osteogenesis imperfecta type Xv and BMD quantitative trait locus 16 (osteoporosis) (Fahiminiya et al., 2013; Keupp et al., 2013; Laine et al., 2013; Pyott et al., 2013; Panigrahi et al., 2018). *CNN1* is a known marker of smooth muscle differentiation, which can bind to actin, tropomyosin, and calmodulin, and is involved in the regulation of smooth muscle contraction activity and cell proliferation (Jiang et al., 1997). It is also expressed in human osteosarcoma cell lines and osteosarcoma tissues (Yamamura et al., 1998). Studies have demonstrated that after knocking out *CNN1*, the degree of ossification in mice increases (Yoshikawa et al., 1998). A recent study found that overexpression of *CNN1* in osteoblasts led to a significant decrease in bone mass at the adult stage by inhibiting osteoblast migration, proliferation, and mineralization, and by promoting osteoclastogenesis (Su et al., 2013). These findings suggest that *CNN1* plays a role in bone-related cells and in the regulation of bone remodeling.

In addition, *NFIX* (Gronostajski, 2000; Adam et al., 2005; Driller et al., 2007; Malan et al., 2010), *CALR* (Klampfl et al., 2013; Nangalia et al., 2013), *FKBP11* (Hanagata et al., 2011; Hanagata and Li, 2011; Tsukamoto et al., 2013), *TRAPPC5* (Sacher et al., 2008; Scrivens et al., 2011), *MAP2K7* (Yamamoto et al., 2002; Svensson et al., 2009), and *RELA* (Frederiksen et al., 2016; Jimi et al., 2019; Yu et al., 2020) have been found to be related to bone health and development in previous studies. Therefore, the above genes are likely to be related to the phenotype of chicken keel, and even affect the development of other bones. Through functional annotation, we identified a new gene related to calmodulin binding (*ENSGALG00000047166*), which is located on chromosome 1. Although this gene has not been reported before, we believe it may be related to bone development. The protein encoded by *PCSK2* is a prohormone processing enzyme involved in insulin and glucagon biosynthesis. Gene polymorphisms are associated with pleiotropic effects on various traits of glucose homeostasis and incident diabetes (Chang et al., 2015). Studies have shown that diabetes is usually accompanied by skeletal fragility (Lecka-Czernik, 2017). Therefore, this gene may indirectly affect bone development.

Enrichment Analysis

We performed GO ontology and KEGG pathway analyses of the converted human homologous genes. Most genes were enriched in the GO term "protein binding," and many genes were also enriched in "cytosol," "nucleoplasm," "nucleus," and "cytoplasm" (corrected p-value < 0.05). We also found 11 significantly enriched KEGG pathways (corrected p-value < 0.05), one of which involves "osteoclast differentiation" and the other is the "MAPK signaling pathway." A large number of studies have reported that the MAPK signaling pathway is involved in osteoblast differentiation (Chen et al., 2012; Yu et al., 2018; Zhao et al., 2018). In KEGG diseases, we found musculoskeletal diseases and congenital malformations. Musculoskeletal diseases often involve multiple tissues (mainly muscles, bones, cartilage, and nerves; Huang and Sowa, 2011). They are major causes of disability worldwide and have been significant in the development of bone and joint diseases (Brooks, 2006). Congenital malformations are defects in the morphogenesis of organs or body districts that are identifiable at birth (Corsello and Giuffre, 2012). In one study, 3,932 newborns were examined for congenital malformations at birth. Among them, the central nervous system (39.5%) was the most commonly involved, followed by the

musculoskeletal system (14.5%; Swain et al., 1994). Through enrichment analysis, We further confirmed that candidate genes screened in this study are likely to be involved in keel development and have an impact on keel phenotype.

CONCLUSION

In summary, we found that KB is a medium heritability trait. In hens with obvious KB, the PTH content was lower and the CT content was higher. The bone strength, BMD, and bone cortical thickness of the humerus and femur of hens with obvious KB were lower than those of KN hens, and there was more adipose tissue in the bone. Our results show that the severity of KB is related to bone strength, BMD, and bone cortical thickness. We conducted a selection elimination analysis based on the differences in keel phenotypes between White Leghorn laying hens and identified 10 important candidate genes that have strong selection signals and are related to bones (*ACP5, WNT1, NFIX, CNN1, CALR, FKBP11, TRAPPC5, MAP2K7, RELA, ENSGALG00000047166*). These genes may be related to changes in the keel phenotype of laying hens. The role of other unreported genes requires further research. In the enrichment analysis, two bone-related pathways, the "MAPK signaling pathway" and "osteoclast differentiation", were enriched, which further verified the reliability of our results. These results may help us to better understand the molecular mechanism of KB in chickens and other birds and provide new insight relevant to the genetic breeding of laying hens.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA 772890.

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Management Committee of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences.

AUTHOR CONTRIBUTIONS

ZZ performed the experiments, analyzed and interpreted the data, and wrote the manuscript. TZ contributed to data analysis. XZ and GZ provided test poultry farms. YJ, LQ, LW, and WY designed the experiments and coordinated the study.YJ and LQ reviewed and revised the manuscript. All authors approved the submitted version.

FUNDING

This work was supported by the Beijing Innovation Team of the Modern Agro-industry Technology Research System (grant number BAIC04-2021).

ACKNOWLEDGMENTS

We sincerely appreciate XZ for access to her chickens and the staff at Hebei Dawu Poultry Breeding Co., Ltd. for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.833132/full#supplementary-material

REFERENCES

Adam, M. P., Hennekam, R. C. M., Keppen, L. D., Bull, M. J., Clericuzio, C. L., Burke, L. W., et al. (2005). Marshall-smith Syndrome: Natural History and Evidence of an Osteochondrodysplasia with Connective Tissue Abnormalities. *Am. J. Med. Genet.* 137A, 117–124. doi:10.1002/ajmg.a.30580

Bilginer, Y., Düzova, A., Topaloğlu, R., Batu, E. D., Boduroğlu, K., Güçer, Ş., et al. (2016). Three Cases of Spondyloenchondrodysplasia (SPENCD) with Systemic Lupus Erythematosus: a Case Series and Review of the Literature. *Lupus* 25, 760–765. doi:10.1177/0961203316629000

Bishop, S. C., Fleming, R. H., McCormack, H. A., Flock, D. K., and Whitehead, C. C. (2000). Inheritance of Bone Characteristics Affecting Osteoporosis in Laying Hens. *Br. Poult. Sci.* 41, 33–40. doi:10.1080/00071660086376

Briggs, T. A., Rice, G. I., Adib, N., Ades, L., Barete, S., Baskar, K., et al. (2016). Spondyloenchondrodysplasia Due to Mutations in ACP5: A Comprehensive Survey. *J. Clin. Immunol.* 36, 220–234. doi:10.1007/s10875-016-0252-y

Brinker, T., Bijma, P., Visscher, J., Rodenburg, T. B., and Ellen, E. D. (2014). Plumage Condition in Laying Hens: Genetic Parameters for Direct and Indirect Effects in Two Purebred Layer Lines. *Genet. Sel. Evol.* 46, 33. doi:10.1186/1297-9686-46-33

Brooks, P. M. (2006). The burden of Musculoskeletal Disease-A Global Perspective. *Clin. Rheumatol.* 25, 778–781. doi:10.1007/s10067-006-0240-3

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-i: Intelligent Prioritization and Exploratory Visualization of Biological Functions for Gene Enrichment Analysis. *Nucleic Acids Res.* 49, W317–W325. doi:10.1093/nar/gkab447

Buckner, G. D., Insko, W. M., J. R., Henry, A. H., and Elizabeth, F. H. (1949). Rate of Growth and Calcification of the Sternum of Male and Female New Hampshire Chickens Having Crooked Keels. *Poult. Sci.* 28, 389–292. doi:10.3382/ps.0280289

Buss, E. G., and Guyer, R. B. (1984). Bone Parameters of Thick and Thin Eggshell Lines of Chickens. *Comp. Biochem. Physiol. A: Physiol.* 78, 449–452. doi:10.1016/0300-9629(84)90576-0

Casey-Trott, T., Heerkens, J. L. T., Petrik, M., Regmi, P., Schrader, L., Toscano, M. J., et al. (2015). Methods for Assessment of Keel Bone Damage in Poultry. *Poult. Sci.* 94, 2339–2350. doi:10.3382/ps/pev223

Casey-Trott, T. M., Guerin, M. T., Sandilands, V., Torrey, S., and Widowski, T. M. (2017). Rearing System Affects Prevalence of Keel-Bone Damage in Laying Hens: a Longitudinal Study of Four Consecutive Flocks. *Poult. Sci.* 96, 2029–2039. doi:10.3382/ps/pex026

Chang, T.-J., Chiu, Y.-F., Sheu, W. H.-H., Shih, K.-C., Hwu, C.-M., Quertermous, T., et al. (2015). Genetic Polymorphisms of PCSK2 Are Associated with Glucose

Homeostasis and Progression to Type 2 Diabetes in a Chinese Population. *Sci. Rep.* 5, 14380. doi:10.1038/srep14380

Chen, G., Deng, C., and Li, Y.-P. (2012). TGF-β and BMP Signaling in Osteoblast Differentiation and Bone Formation. *Int. J. Biol. Sci.* 8, 272–288. doi:10.7150/ijbs.2929

Claessens, L. P. A. M. (2009). The Skeletal Kinematics of Lung Ventilation in Three Basal Bird Taxa (Emu, Tinamou, and guinea Fowl). *J. Exp. Zool.* 311A, 586–599. doi:10.1002/jez.501

Corsello, G., and Giuffrè, M. (2012). Congenital Malformations. *J. Maternal-Fetal Neonatal Med.* 25 (Suppl. 1), 25–29. doi:10.3109/14767058.2012.664943

Daumer, K. M., Tufan, A. C., and Tuan, R. S. (2004). Long-term *In Vitro* Analysis of Limb Cartilage Development: Involvement of Wnt Signaling. *J. Cel. Biochem.* 93, 526–541. doi:10.1002/jcb.20190

Demontiero, O., Vidal, C., and Duque, G. (2012). Aging and Bone Loss: New Insights for the Clinician. *Ther. Adv. Musculoskelet.* 4, 61–76. doi:10.1177/1759720X11430858

Di Iorgi, N., Mo, A. O., Grimm, K., Wren, T. A. L., Dorey, F., and Gilsanz, V. (2010). Bone Acquisition in Healthy Young Females Is Reciprocally Related to Marrow Adiposity. *J. Clin. Endocrinol. Metab.* 95, 2977–2982. doi:10.1210/jc.2009-2336

Di Iorgi, N., Rosol, M., Mittelman, S. D., and Gilsanz, V. (2008). Reciprocal relation between marrow adiposity and the amount of bone in the axial and appendicular skeleton of young adults. *J. Clin. Endocr. Metab.* 93, 2281–2286. doi:10.1210/jc.2007-2691

Dominguez-Gasca, N., Benavides-Reyes, C., Sánchez-Rodríguez, E., and Rodríguez-Navarro, A. B. (2019). Changes in avian cortical and medullary bone mineral composition and organization during acid-induced demineralization. *ejm* 31, 209–216. doi:10.1127/ejm/2019/0031-2826

Donaldson, C. J., Ball, M. E. E., and O'Connell, N. E. (2012). Aerial Perches and Free-Range Laying Hens: the Effect of Access to Aerial Perches and of Individual Bird Parameters on Keel Bone Injuries in Commercial Free-Range Laying Hens. *Poult. Sci.* 91, 304–315. doi:10.3382/ps.2011-01774

Driller, K., Pagenstecher, A., Uhl, M., Omran, H., Berlis, A., Gründer, A., et al. (2007). Nuclear Factor I X Deficiency Causes Brain Malformation and Severe Skeletal Defects. *Mol. Cel. Biol.* 27, 3855–3867. doi:10.1128/MCB.02293-06

Edmonson, A. J., Lean, I. J., Weaver, L. D., Farver, T., and Webster, G. (1989). A Body Condition Scoring Chart for Holstein Dairy Cows. *J. Dairy Sci.* 72, 68–78. doi:10.3168/jds.S0022-0302(89)79081-0

Fahiminiya, S., Majewski, J., Mort, J., Moffatt, P., Glorieux, F. H., and Rauch, F. (2013). Mutations in WNT1 Are a Cause of Osteogenesis Imperfecta. *J. Med. Genet.* 50, 345–348. doi:10.1136/jmedgenet-2013-101567

Fleming, R. H., McCormack, H. A., McTeir, L., and Whitehead, C. C. (2004). Incidence, Pathology and Prevention of Keel Bone Deformities in the Laying Hen. *Br. Poult. Sc.* 45, 320–330. doi:10.1080/00071660410001730815

Frederiksen, A. L., Larsen, M. J., Brusgaard, K., Novack, D. V., Knudsen, P. J. T., Schrøder, H. D., et al. (2016). Neonatal High Bone Mass with First Mutation of the NF-Kb Complex: Heterozygous De Novo Missense (p.Asp512Ser) RELA(Rela/p65). *J. Bone Miner. Res.* 31, 163–172. doi:10.1002/jbmr.2590

Gebhardt-Henrich, S., and Fröhlich, E. (2015). Early Onset of Laying and Bumblefoot Favor Keel Bone Fractures. *Animals* 5, 1192–1206. doi:10.3390/ani5040406

Gebhardt-Henrich, S. G., Pfulg, A., Fröhlich, E. K. F., Käppeli, S., Guggisberg, D., Liesegang, A., et al. (2017). Limited Associations between Keel Bone Damage and Bone Properties Measured with Computer Tomography, Three-Point Bending Test, and Analysis of Minerals in Swiss Laying Hens. *Front. Vet. Sci.* 4, 128. doi:10.3389/fvets.2017.00128

Girschick, H., Wolf, C., Morbach, H., Hertzberg, C., and Lee-Kirsch, M. A. (2015). Severe Immune Dysregulation with Neurological Impairment and Minor Bone Changes in a Child with Spondyloenchondrodysplasia Due to Two Novel Mutations in the ACP5 Gene. *Pediatr. Rheumatol.* 13, 37. doi:10.1186/s12969-015-0035-7

Goltzman, D. (2018). Physiology of Parathyroid Hormone. *Endocrinol. Metab. Clin. North America* 47, 743–758. doi:10.1016/j.ecl.2018.07.003

Gronostajski, R. M. (2000). Roles of the NFI/CTF Gene Family in Transcription and Development. *Gene* 249, 31–45. doi:10.1016/s0378-1119(00)00140-2

Guo, J., Qu, L., Dou, T.-C., Shen, M.-M., Hu, Y.-P., Ma, M., et al. (2020). Genome-wide Association Study Provides Insights into the Genetic Architecture of Bone Size and Mass in Chickens. *Genome* 63, 133–143. doi:10.1139/gen-2019-0022

Guo, J., Sun, C., Qu, L., Shen, M., Dou, T., Ma, M., et al. (2017). Genetic Architecture of Bone Quality Variation in Layer Chickens Revealed by a Genome-wide Association Study. *Sci. Rep.* 7, 45317. doi:10.1038/srep45317

Hanagata, N., Li, X., Morita, H., Takemura, T., Li, J., and Minowa, T. (2011). Characterization of the Osteoblast-specific Transmembrane Protein IFITM5 and Analysis of IFITM5-Deficient Mice. *J. Bone Miner. Metab.* 29, 279–290. doi:10.1007/s00774-010-0221-0

Hanagata, N., and Li, X. (2011). Osteoblast-enriched Membrane Protein IFITM5 Regulates the Association of CD9 with an FKBP11-CD81-FPRP Complex and Stimulates Expression of Interferon-Induced Genes. *Biochem. Biophysical Res. Commun.* 409, 378–384. doi:10.1016/j.bbrc.2011.04.136

Hartmann, C., and Tabin, C. J. (2000). Dual Roles of Wnt Signaling during Chondrogenesis in the Chicken Limb. *Development* 127, 3141–3159. doi:10.1242/dev.127.14.3141

Hayman, A. R., Jones, S. J., Boyde, A., Foster, D., Colledge, W. H., Carlton, M. B., et al. (1996). Mice Lacking Tartrate-Resistant Acid Phosphatase (Acp 5) Have Disrupted Endochondral Ossification and Mild Osteopetrosis. *Development* 122, 3151–3162. doi:10.1242/dev.122.10.3151

Hozumi, A., Osaki, M., Goto, H., Sakamoto, K., Inokuchi, S., and Shindo, H. (2009). Bone Marrow Adipocytes Support Dexamethasone-Induced Osteoclast Differentiation. *Biochem. Biophysical Res. Commun.* 382, 780–784. doi:10.1016/j.bbrc.2009.03.111

Huang, W., and Sowa, G. (2011). Biomarker Development for Musculoskeletal Diseases. *PM&R* 3, S39–S44. doi:10.1016/j.pmrj.2011.04.023

Hudson, H. A., Britton, W. M., Rowland, G. N., and Buhr, R. J. (1993). Histomorphometric Bone Properties of Sexually Immature and Mature White Leghorn Hens with Evaluation of Fluorochrome Injection on Egg Production Traits. *Poult. Sci.* 72, 1537–1547. doi:10.3382/ps.0721537

Jahejo, A. R., Niu, S., Zhang, D., Ning, G.-b., Khan, A., Mangi, R. A., et al. (2019). Transcriptome Analysis of MAPK Signaling Pathway and Associated Genes to Angiogenesis in Chicken Erythrocytes on Response to Thiram-Induced Tibial Lesions. *Res. Vet. Sci.* 127, 65–75. doi:10.1016/j.rvsc.2019.10.013

Jiang, Z., Grange, R. W., Walsh, M. P., and Kamm, K. E. (1997). Adenovirus-mediated Transfer of the Smooth Muscle Cell Calponin Gene Inhibits Proliferation of Smooth Muscle Cells and Fibroblasts. *FEBS Lett.* 413, 441–445. doi:10.1016/s0014-5793(97)00944-7

Jimi, E., Huang, F., and Nakatomi, C. (2019). NF-κB Signaling Regulates Physiological and Pathological Chondrogenesis. *Ijms* 20, 6275. doi:10.3390/ijms20246275

Joeng, K. S., Lee, Y.-C., Lim, J., Chen, Y., Jiang, M.-M., Munivez, E., et al. (2017). Osteocyte-specific WNT1 Regulates Osteoblast Function during Bone Homeostasis. *J. Clin. Invest.* 127, 2678–2688. doi:10.1172/JCI92617

Kebreab, E., France, J., Kwakkel, R. P., Leeson, S., Kuhi, H. D., and Dijkstra, J. (2009). Development and Evaluation of a Dynamic Model of Calcium and Phosphorus Flows in Layers. *Poult. Sci.* 88, 680–689. doi:10.3382/ps.2008-00157

Kerschnitzki, M., Zander, T., Zaslansky, P., Fratzl, P., Shahar, R., and Wagermaier, W. (2014). Rapid Alterations of Avian Medullary Bone Material during the Daily Egg-Laying Cycle. *Bone* 69, 109–117. doi:10.1016/j.bone.2014.08.019

Keupp, K., Beleggia, F., Kayserili, H., Barnes, A. M., Steiner, M., Semler, O., et al. (2013). Mutations in WNT1 Cause Different Forms of Bone Fragility. *Am. J. Hum. Genet.* 92, 565–574. doi:10.1016/j.ajhg.2013.02.010

Kijas, J. W. (2014). Haplotype-based Analysis of Selective Sweeps in Sheep. *Genome* 57, 433–437. doi:10.1139/gen-2014-0049

Klampfl, T., Gisslinger, H., Harutyunyan, A. S., Nivarthi, H., Rumi, E., Milosevic, J. D., et al. (2013). Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *N. Engl. J. Med.* 369, 2379–2390. doi:10.1056/NEJMoa1311347

Laine, C. M., Joeng, K. S., Campeau, P. M., Kiviranta, R., Tarkkonen, K., Grover, M., et al. (2013). WNT1 Mutations in Early-Onset Osteoporosis and Osteogenesis Imperfecta. *N. Engl. J. Med.* 368, 1809–1816. doi:10.1056/NEJMoa1215458

Lausch, E., Janecke, A., Bros, M., Trojandt, S., Alanay, Y., De Laet, C., et al. (2011). Genetic Deficiency of Tartrate-Resistant Acid Phosphatase Associated with Skeletal Dysplasia, Cerebral Calcifications and Autoimmunity. *Nat. Genet.* 43, 132–137. doi:10.1038/ng.749

Lay, D. C., Jr., Fulton, R. M., Hester, P. Y., Karcher, D. M., Kjaer, J. B., Mench, J. A., et al. (2011). Hen Welfare in Different Housing Systems. *Poult. Sci.* 90, 278–294. doi:10.3382/ps.2010-00962

Lecka-Czernik, B. (2017). Diabetes, Bone and Glucose-Lowering Agents: Basic Biology. *Diabetologia* 60, 1163–1169. doi:10.1007/s00125-017-4269-4

Li, X., Zhang, D., and Bryden, W. L. (2017). Calcium and Phosphorus Metabolism and Nutrition of Poultry: Are Current Diets Formulated in Excess? *Anim. Prod. Sci.* 57, 2304–2310. doi:10.1071/an17389

Li, Y. D., Liu, X., Li, Z. W., Wang, W. J., Li, Y. M., Cao, Z. P., et al. (2021). A Combination of Genome-wide Association Study and Selection Signature Analysis Dissects the Genetic Architecture Underlying Bone Traits in Chickens. *Animal* 15, 100322. doi:10.1016/j.animal.2021.100322

Lu, Y., Ren, X., Wang, Y., Bardai, G., Sturm, M., Dai, Y., et al. (2018). Novel WNT1 Mutations in Children with Osteogenesis Imperfecta: Clinical and Functional Characterization. *Bone* 114, 144–149. doi:10.1016/j.bone.2018.06.018

Madsen, P., Sørensen, P., Su, G., Damgaard, L. H., Thomsen, H., and Labouriau, R. (2006). *DMU - A Package for Analyzing Multivariate Mixed Models*. Minas Gerais: Instituto Prociência. ID - 20063170093, 11–27.

Maeda, K., Kobayashi, Y., Koide, M., Uehara, S., Okamoto, M., Ishihara, A., et al. (2019). The Regulation of Bone Metabolism and Disorders by Wnt Signaling. *Ijms* 20, 5525. doi:10.3390/ijms20225525

Malan, V., Rajan, D., Thomas, S., Shaw, A. C., Louis Dit Picard, H., Layet, V., et al. (2010). Distinct Effects of Allelic NFIX Mutations on Nonsense-Mediated mRNA Decay Engender Either a Sotos-like or a Marshall-Smith Syndrome. *Am. J. Hum. Genet.* 87, 189–198. doi:10.1016/j.ajhg.2010.07.001

Martin, R. B., and Zissimos, S. L. (1991). Relationships between Marrow Fat and Bone Turnover in Ovariectomized and Intact Rats. *Bone* 12, 123–131. doi:10.1016/8756-3282(91)90011-7

Maurin, A. C., Chavassieux, P. M., Vericel, E., and Meunier, P. J. (2002). Role of Polyunsaturated Fatty Acids in the Inhibitory Effect of Human Adipocytes on Osteoblastic Proliferation. *Bone* 31, 260–266. doi:10.1016/s8756-3282(02)00805-0

Mueller, S., Kreuzer, M., Siegrist, M., Mannale, K., Messikommer, R. E., and Gangnat, I. D. M. (2018). Carcass and Meat Quality of Dual-Purpose Chickens (Lohmann Dual, Belgian Malines, Schweizerhuhn) in Comparison to Broiler and Layer Chicken Types. *Poult. Sci.* 97, 3325–3336. doi:10.3382/ps/pey172

Mueller, W. J., Schraer, R., and Scharer, H. (1964). Calcium Metabolism and Skeletal Dynamics of Laying Pullets. *J. Nutr.* 84, 20–26. doi:10.1093/jn/84.1.20

Nangalia, J., Massie, C. E., Baxter, E. J., Nice, F. L., Gundem, G., Wedge, D. C., et al. (2013). Somatic CALR Mutations in Myeloproliferative Neoplasms with Nonmutated JAK2. *N. Engl. J. Med.* 369, 2391–2405. doi:10.1056/NEJMoa1312542

Naot, D., Musson, D. S., and Cornish, J. (2019). The Activity of Peptides of the Calcitonin Family in Bone. *Physiol. Rev.* 99, 781–805. doi:10.1152/physrev.00066.2017

Nasr, M. A. F., Browne, W. J., Caplen, G., Hothersall, B., Murrell, J. C., and Nicol, C. J. (2013). Positive Affective State Induced by Opioid Analgesia in Laying Hens with Bone Fractures. *Appl. Anim. Behav. Sci.* 147, 127–131. doi:10.1016/j.applanim.2013.04.015

Nasr, M. A. F., Nicol, C. J., and Murrell, J. C. (2012b). Do laying Hens with Keel Bone Fractures Experience Pain? *PLoS One* 7, e42420. doi:10.1371/journal.pone.0042420

Nasr, M., Murrell, J., Wilkins, L., and Nicol, C. (2012a). The Effect of Keel Fractures on Egg-Production Parameters, Mobility and Behaviour in Individual Laying Hens. *Anim. Welf* 21, 127–135. doi:10.7120/096272812799129376

Nys, Y., and Le Roy, N. (2018). *Calcium Homeostasis and Eggshell Biomineralization in Female Chicken*. London: Academic Press.

Ødegård, J., Meuwissen, T. H., Heringstad, B., and Madsen, P. (2010). A Simple Algorithm to Estimate Genetic Variance in an Animal Threshold Model Using Bayesian Inference. *Genet. Sel. Evol.* 42, 29. doi:10.1186/1297-9686-42-29

Odihambo Mumma, J., Thaxton, J. P., Vizzier-Thaxton, Y., and Dodson, W. L. (2006). Physiological Stress in Laying Hens. *Poult. Sci.* 85, 761–769. doi:10.1093/ps/85.4.761

Owen, M., and Friedenstein, A. J. (1988). Stromal Stem Cells: Marrow-Derived Osteogenic Precursors. *Ciba Found. Symp.* 136, 42–60. doi:10.1002/9780470513637.ch4

Panigrahi, I., Didel, S., Kirpal, H., Bellampalli, R., Miyanath, S., Mullapudi, N., et al. (2018). Novel Mutation in a Family with WNT1 -related Osteoporosis. *Eur. J. Med. Genet.* 61, 369–371. doi:10.1016/j.ejmg.2018.01.017

Pyott, S. M., Tran, T. T., Leistritz, D. F., Pepin, M. G., Mendelsohn, N. J., Temme, R. T., et al. (2013). WNT1 Mutations in Families Affected by Moderately Severe and Progressive Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 92, 590–597. doi:10.1016/j.ajhg.2013.02.009

Ramesh, J., Parthasarathy, L. K., Janckila, A. J., Begum, F., Murugan, R., Murthy, B. P. S. S., et al. (2020). Characterisation of ACP5 Missense Mutations Encoding Tartrate-Resistant Acid Phosphatase Associated with Spondyloenchondrodysplasia. *PLoS One* 15, e0230052. doi:10.1371/journal.pone.0230052

Riber, A. B., Casey-Trott, T. M., and Herskin, M. S. (2018). The Influence of Keel Bone Damage on Welfare of Laying Hens. *Front. Vet. Sci.* 5, 6. doi:10.3389/fvets.2018.00006

Roberts, J. R. (2004). Factors Affecting Egg Internal Quality and Egg Shell Quality in Laying Hens. *J. Poult. Sci.* 41, 161–177. doi:10.2141/jpsa.41.161

Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., et al. (2018). Revealing the Selection History of Adaptive Loci Using Genome-wide Scans for Selection: an Example from Domestic Sheep. *BMC Genomics* 19, 71. doi:10.1186/s12864-018-4447-x

Rodriguez-Navarro, A. B., McCormack, H. M., Fleming, R. H., Alvarez-Lloret, P., Romero-Pastor, J., Dominguez-Gasca, N., et al. (2018). Influence of Physical Activity on Tibial Bone Material Properties in Laying Hens. *J. Struct. Biol.* 201, 36–45. doi:10.1016/j.jsb.2017.10.011

Rosen, C. J., and Bouxsein, M. L. (2006). Mechanisms of Disease: Is Osteoporosis the Obesity of Bone? *Nat. Rev. Rheumatol.* 2, 35–43. doi:10.1038/ncprheum0070

Sabatakos, G., Sims, N. A., Chen, J., Aoki, K., Kelz, M. B., Amling, M., et al. (2000). Overexpression of ΔFosB Transcription Factor(s) Increases Bone Formation and Inhibits Adipogenesis. *Nat. Med.* 6, 985–990. doi:10.1038/79683

Sacher, M., Kim, Y.-G., Lavie, A., Oh, B.-H., and Segev, N. (2008). The TRAPP Complex: Insights into its Architecture and Function. *Traffic* 9, 2032–2042. doi:10.1111/j.1600-0854.2008.00833.x

Scrivens, P. J., Noueihed, B., Shahrzad, N., Hul, S., Brunet, S., and Sacher, M. (2011). C4orf41 and TTC-15 Are Mammalian TRAPP Components with a Role at an Early Stage in ER-To-Golgi Trafficking. *MBoC* 22, 2083–2093. doi:10.1091/mbc.E10-11-0873

Shen, W., Chen, J., Gantz, M., Punyanitya, M., Heymsfield, S. B., Gallagher, D., et al. (2012a). MRI-measured Pelvic Bone Marrow Adipose Tissue Is Inversely Related to DXA-Measured Bone mineral in Younger and Older Adults. *Eur. J. Clin. Nutr.* 66, 983–988. doi:10.1038/ejcn.2012.35

Shen, W., Chen, J., Punyanitya, M., Shapses, S., Heshka, S., and Heymsfield, S. B. (2007). MRI-measured Bone Marrow Adipose Tissue Is Inversely Related to DXA-Measured Bone mineral in Caucasian Women. *Osteoporos. Int.* 18, 641–647. doi:10.1007/s00198-006-0285-9

Shen, W., Scherzer, R., Gantz, M., Chen, J., Punyanitya, M., Lewis, C. E., et al. (2012b). Relationship between MRI-Measured Bone Marrow Adipose Tissue and Hip and Spine Bone mineral Density in African-American and Caucasian Participants: the CARDIA Study. *J. Clin. Endocrinol. Metab.* 97, 1337–1346. doi:10.1210/jc.2011-2605

Su, N., Chen, M., Chen, S., Li, C., Xie, Y., Zhu, Y., et al. (2013). Overexpression of H1 Calponin in Osteoblast Lineage Cells Leads to a Decrease in Bone Mass by Disrupting Osteoblast Function and Promoting Osteoclast Formation. *J. Bone Miner. Res.* 28, 660–671. doi:10.1002/jbmr.1778

Svensson, C. I., Inoue, T., Hammaker, D., Fukushima, A., Papa, S., Franzoso, G., et al. (2009). Gadd45β Deficiency in Rheumatoid Arthritis: Enhanced Synovitis through JNK Signaling. *Arthritis Rheum.* 60, 3229–3240. doi:10.1002/art.24887

Swain, S., Agrawal, A., and Bhatia, B. D. (1994). Congenital Malformations at Birth. *Indian Pediatr.* 31, 1187–1191.

Toscano, M. J., Dunn, I. C., Christensen, J.-P., Petow, S., Kittelsen, K., and Ulrich, R. (2020). Explanations for Keel Bone Fractures in Laying Hens: Are There Explanations in Addition to Elevated Egg Production? *Poult. Sci.* 99, 4183–4194. doi:10.1016/j.psj.2020.05.035

Tsukamoto, T., Li, X., Morita, H., Minowa, T., Aizawa, T., Hanagata, N., et al. (2013). Role of S-Palmitoylation on IFITM5 for the Interaction with FKBP11 in Osteoblast Cells. *PLoS One* 8, e75831. doi:10.1371/journal.pone.0075831

Wang, F., Rummukainen, P., Heino, T. J., and Kiviranta, R. (2021). Osteoblastic Wnt1 Regulates Periosteal Bone Formation in Adult Mice. *Bone* 143, 115754. doi:10.1016/j.bone.2020.115754

Wehrli, F. W., Hopkins, J. A., Hwang, S. N., Song, H. K., Snyder, P. J., and Haddad, J. G. (2000). Cross-sectional Study of Osteopenia with Quantitative MR Imaging and Bone Densitometry. *Radiology* 217, 527–538. doi:10.1148/radiology.217.2.r00nv20527

Wei, H., Bi, Y., Xin, H., Pan, L., Liu, R., Li, X., et al. (2020). Keel Fracture Changed the Behavior and Reduced the Welfare, Production Performance, and Egg Quality in Laying Hens Housed Individually in Furnished Cages. *Poult. Sci.* 99, 3334–3342. doi:10.1016/j.psj.2020.04.001

Whitehead, C. C. (2004). Overview of Bone Biology in the Egg-Laying Hen. *Poult. Sci.* 83, 193–199. doi:10.1093/ps/83.2.193

Yamamoto, A., Miyazaki, T., Kadono, Y., Takayanagi, H., Miura, T., Nishina, H., et al. (2002). Possible Involvement of IκB Kinase 2 and MKK7 in Osteoclastogenesis Induced by Receptor Activator of Nuclear Factor κB Ligand. *J. Bone Miner. Res.* 17, 612–621. doi:10.1359/jbmr.2002.17.4.612

Yamamura, H., Yoshikawa, H., Tatsuta, M., Akedo, H., and Takahashi, K. (1998). Expression of the Smooth Muscle Calponin Gene in Human Osteosarcoma and its Possible Association with Prognosis. *Int. J. Cancer* 79, 245–250. doi:10.1002/(sici)1097-0215(19980619)79:3<245::aid-ijc6>3.0.co;2-p

Yoshikawa, H., Taniguchi, S. i., Yamamura, H., Mori, S., Sugimoto, M., Miyado, K., et al. (1998). Mice Lacking Smooth Muscle Calponin Display Increased Bone Formation that Is Associated with Enhancement of Bone Morphogenetic Protein Responses. *Genes to Cells* 3, 685–695. doi:10.1046/j.1365-2443.1998.00214.x

Yu, S., Li, P., Li, B., Miao, D., and Deng, Q. (2020). RelA Promotes Proliferation but Inhibits Osteogenic and Chondrogenic Differentiation of Mesenchymal Stem Cells. *FEBS Lett.* 594, 1368–1378. doi:10.1002/1873-3468.13739

Yu, X., Quan, J., Long, W., Chen, H., Wang, R., Guo, J., et al. (2018). LL-37 Inhibits LPS-Induced Inflammation and Stimulates the Osteogenic Differentiation of BMSCs via P2X7 Receptor and MAPK Signaling Pathway. *Exp. Cel Res.* 372, 178–187. doi:10.1016/j.yexcr.2018.09.024

Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt Signaling in Cancer. *Oncogene* 36, 1461–1473. doi:10.1038/onc.2016.304

Zhao, P., Xiao, L., Peng, J., Qian, Y. Q., and Huang, C. C. (2018). Exosomes Derived from Bone Marrow Mesenchymal Stem Cells Improve Osteoporosis through Promoting Osteoblast Proliferation via MAPK Pathway. *Eur. Rev. Med. Pharmacol. Sci.* 22, 3962–3970. doi:10.26355/eurrev_201806_15280

Zheng, M.-H., Li, F.-X. -Z., Xu, F., Lin, X., Wang, Y., Xu, Q.-S., et al. (2020). The Interplay between the Renin-Angiotensin-Aldosterone System and Parathyroid Hormone. *Front. Endocrinol.* 11, 539. doi:10.3389/fendo.2020.00539

Zhu, T., Zhang, T.-Y., Wen, J., Zhao, X., Chen, Y., Jia, Y., et al. (2020). The Genetic Architecture of the Chickens Dropping Moisture by Genetic Parameter Estimation and Genome-wide Association. *Front. Genet.* 11, 806. doi:10.3389/fgene.2020.00806

Conflict of Interest: Authors XZ and GZ were employed by the company Hebei Dawu Poultry Breeding Co., Ltd.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/fgene.2022.784663

Extensive Variation in Gene Expression is Revealed in 13 Fertility-Related Genes Using RNA-Seq, ISO-Seq, and CAGE-Seq From Brahman Cattle

Elizabeth M. Ross[1*†], *Hari Sanjana*[1†], *Loan T. Nguyen*[1], *YuanYuan Cheng*[2], *Stephen S. Moore*[1] *and Ben J. Hayes*[1]

[1]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia, [2]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Yang Zhou,
Huazhong Agricultural University, China

Reviewed by:
Tara G. McDaneld, Agricultural Research Service (USDA), United States
Haibo Liu, University of Massachusetts Medical School, United States

***Correspondence:**
Elizabeth M. Ross
e.ross@uq.edu.au

†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: *28 September 2021*
Accepted: *24 January 2022*
Published: *25 March 2022*

Citation:
Ross EM, Sanjana H, Nguyen LT, Cheng Y, Moore SS and Hayes BJ (2022) Extensive Variation in Gene Expression is Revealed in 13 Fertility-Related Genes Using RNA-Seq, ISO-Seq, and CAGE-Seq From Brahman Cattle. Front. Genet. 13:784663. doi: 10.3389/fgene.2022.784663

Fertility is a key driver of economic profitability in cattle production. A number of studies have identified genes associated with fertility using genome wide association studies and differential gene expression analysis; however, the genes themselves are poorly characterized in cattle. Here, we selected 13 genes from the literature which have previously been shown to have strong evidence for an association with fertility in Brahman cattle (*Bos taurus indicus*) or closely related breeds. We examine the expression variation of the 13 genes that are associated with cattle fertility using RNA-seq, CAGE-seq, and ISO-seq data from 11 different tissue samples from an adult Brahman cow and a Brahman fetus. Tissues examined include blood, liver, lung, kidney, muscle, spleen, ovary, and uterus from the cow and liver and lung from the fetus. The analysis revealed several novel isoforms, including seven from *SERPINA7*. The use of three expression characterization methodologies (5′ cap selected ISO-seq, CAGE-seq, and RNA-seq) allowed the identification of isoforms that varied in their length of 5′ and 3′ untranslated regions, variation otherwise undetectable (collapsed as degraded RNA) in generic isoform identification pipelines. The combinations of different sequencing technologies allowed us to overcome the limitations of relatively low sequence depth in the ISO-seq data. The lower sequence depth of the ISO-seq data was also reflected in the lack of observed expression of some genes that were observed in the CAGE-seq and RNA-seq data from the same tissue. We identified allele specific expression that was tissue-specific in *AR*, *IGF1*, *SOX9*, *STAT3*, and *TAF9B*. Finally, we characterized an exon of *TAF9B* as partially nested within the neighboring gene phosphoglycerate kinase 1. As this study only examined two animals, even more transcriptional variation may be present in a genetically diverse population. This analysis reveals the large amount of transcriptional variation within mammalian fertility genes and illuminates the fact that the transcriptional landscape cannot be fully characterized using a single technology alone.

Keywords: fertility, cattle, gene expression, transcriptomics, variation, isoforms, androgen receptor, RNA sequencing

INTRODUCTION

In tropical regions, *Bos taurus indicus* and crosses between *Bos taurus indicus* and *Bos taurus taurus* are extensively used as they are more resistant to heat stress, diseases, and ticks. Brahman cattle are a *Bos taurus indicus* breed extensively raised in tropical regions, including Northern Australia, Brazil, South Asia, and North America. In tropical beef production, fertility is a major driver of profitability; fertility levels can make the difference between a profitable and non-profitable enterprise. Despite the importance of fertility in tropical beef cattle, relatively little is known about the actual genes which contribute toward the genetic variation of the trait.

The FAANG (Functional Annotation of Animal Genomes) data types aim to characterize genes within the genomes of important animal species and breeds, eventually leading to an understanding of how genetic variation translates to phenotypic variation. FAANG data include but are not limited to 1) RNA-seq (Ozsolak and Milos, 2011), which uses short-read sequencing to quantify gene expression, as well as providing some information on gene structure through the presence of intron spanning reads (Wang Z. et al., 2009); 2) ISO-seq (Gonzalez-Garay, 2016), which uses long-read sequencing technology to identify isoforms and provide some information on gene abundance; and 3) CAGE-seq (cap analysis gene expression sequencing), which uses short-read sequencing and captures the 5′ end guanosine caps of eukaryotic mRNAs to identify transcription start sites (Takahashi et al., 2012a; Takahashi et al., 2012b; Salavati et al., 2020). Through the use of these technologies in combination, it is possible to characterize the structure and abundance of genes and identify potential mechanisms, such as allele-specific expression, that can lead to biological diversity.

A number of genes relevant to Brahman fertility have been identified using genome-wide association studies (Fortes et al., 2012a; Minten et al., 2013; Mota et al., 2017; Müller et al., 2017) and expression studies (Beerda et al., 2008; Moore et al., 2016; Dias et al., 2017; Nguyen et al., 2017; Nguyen et al., 2018; Nguyen et al., 2019). But within the Brahman breed, very little work has been conducted to characterize these critical genes. Here, we apply data from three gene expression datasets generated from a Brahman cow and fetus obtained from the same animal to characterize the transcriptional variation present within these genes within the Brahman breed of cattle.

MATERIALS AND METHODS

Overview

In this study, genes associated with fertility traits in the literature were characterized in two Brahman cattle. Gene sequences from *Bos taurus* annotated genome (Rosen et al., 2020) were located within the Brahman genome (Ross et al., 2022) using BLASTn. CAGE-seq, ISO-seq, and RNA-seq data from 10 different tissues were mapped to these genes. The tissues used were from the same animal as was used to generate the Brahman genome assembly, and tissues from a female fetus she was carrying at the time of slaughter were also taken. The expression characteristics from each of these datatypes were then examined to characterize the expression variation within each gene.

Sample Collection

The tissues were obtained post commercial slaughter of an Australian Brahman cow from a commercial abattoir. Spleen, longissiumus dorsi muscle, thyroid, ovary, kidney, uterus, lungs, blood, and liver tissues were obtained from the Brahman cow, and lung and liver tissues were obtained from its developing fetus. The tissue samples were collected post commercial slaughter and immediately snap-frozen in liquid nitrogen. The samples were transferred on dry ice and then stored at −80°C until processing.

Locating Genes in Brahman Genome

Genes that have been previously associated with fertility traits in cattle were identified in the literature. Evidence for an association between the gene and variation in fertility traits included close proximity to a genome-wide association peak or a significant result in a differential expression RNA analysis and other public information, such as a fertility-related phenotype in other species (**Table 1**).

The coding sequence (CDS) of each target gene was downloaded from the *Bos taurus* gene database of National Centre for Biotechnology Information (NCBI) (Rosen et al., 2020). *Bos taurus* gene sequences were aligned to the genome assembly of the Brahman animal (Ross et al., 2022) using BLASTn (Zhang et al., 2000; National Center for Biotechnology Information, 2008), with an e-value cutoff of 10^{-10}.

Short-Read Whole Genome Sequencing

To identify and confirm the genomic sequence of the genes and to identify heterozygous loci, short-read data from both the Brahman cow and the fetus that were sequenced on the Novaseq6000 S4 flow-cell on a 2 × 150 bp paired-end run were aligned to the Brahman genome assembly (Ross et al., 2022).

To remove low-quality data and adapters, the sequences were quality-trimmed before analysis. The raw sequence data were quality-trimmed using the program QUADTrim (Robinson and Ross, 2019) using the options "*-m 10*" to direct QUADTrim to perform quality trim and N-base filter; "*-g*" to remove the guanosine tail, an error that often results from the optics of the NovaSeq6000; and "*-d bulls*" to specify the pre-set trimming filters specified in the 1000 Bull Genome Project (Hayes and Daetwyler, 2019).

Additionally, to identify genome-wide SNP, the 1000 Bull genomes pipeline (23) was followed. Briefly, reads were aligned to the ARS1.2 Bos taurus genome assembly. Alignments were quality-filtered, and SNPs were called. Loci that were called as the homozygous reference (*Bos taurus*) were removed. SNPs which were homozygous alternative (Ho) and heterozygous (He) were counted. Using the assumption that the whole genome contained 2.7 Gbp, the percentage divergence (D) between the whole genome at the haplotype level was calculated as:

TABLE 1 | Genes (identified from literature) as involved in Brahman fertility.

Full Gene Name (Abbreviate gene name)	Function	Evidence for association with fertility in Brahman
Androgen receptor (*AR*)	Member of nuclear receptor superfamily (Fortes et al., 2012b); regulates male fertility *via* regulation of androgen (Brinkmann et al., 1999; Wang R.-S. et al., 2009); interruption of gene function may cause spermatogenesis impairment; feminine character development in males (humans); cancers of prostate, ovary, or breast (MacLean et al., 1995; Dowsing et al., 1999).	Positional candidate in GWAS affecting male and female traits (Lyons et al., 2014).
Insulin-like growth factor 1 (*IGF1*)	Aids in cell growth, differentiation, and transformation (Le Reith, 1997; Yilmaz et al., 2006), regulates release of GnRH which affects age of onset of puberty, conception rate, and maintenance of pregnancy in mammals (Simmen et al., 1993; Wilson, 1998; Velazquez et al., 2008); *IGF1* concentration positively regulates scrotal circumference, motile sperm quantity, and calving rate (Yilmaz et al., 2004; Fortes et al., 2012b).	Positional candidate affecting male and female traits in GWAS including related cattle breeds (Fortes et al., 2013a; Mota et al., 2017).
Inhibin subunit alpha (*INHA*)	Codes for α subunit of inhibin protein (Stelzer et al., 2016); protein complex with α and β subunit negatively regulates secretion of FSH (Richards and Pangas, 2010; Kaneko, 2016; Ulloa-Aguirre et al., 2018); regulates variability of inhibin before puberty (Fortes et al., 2012b).	Candidate gene for GWAS (Fortes et al., 2012b) in males and secretion related to ovulation in females (Burns et al., 2005).
Proenkephalin (*PENK*)	Codes for the neurotransmitters methionine enkephalin and leucine enkephalin *via* proteolytic cleavage (Takahashi, 2016); negatively regulates GnRH *via* modulation of progesterone release (Taylor et al., 2007); negatively regulates LH expression (Malven, 1995)	Biological candidate for GWAS affecting male and female traits (Fortes et al., 2012b) and SNP detection through RNA-seq study (Dias et al., 2017).
Pleiomorphic adenoma gene 1 (*PLAG1*)	Family of zinc finger transcription factors (Juma et al., 2016) influences the stature of cattle which negatively affects fertility (Karim et al., 2011), height of hip, food intake, calving ease, weight at birth, further weight gain, and body size (Snelling et al., 2010; Karim et al., 2011; Pausch et al., 2011; Littlejohn et al., 2012; Fortes et al., 2013b; Juma et al., 2016; Bouwman et al., 2018).	Haplotype analysis (Utsunomiya et al., 2017) and candidate gene for GWAS affecting male and female traits (Littlejohn et al., 2012; Fortes et al., 2013b; Mota et al., 2017).
Ribosomal protein S20 (*RPS20*)	Family of S10P ribosomal proteins (O'Leary et al., 2016) pleiotropically affects body size (McGowan et al., 2008); candidate gene for calving ease and puberty (Pausch et al., 2011; Fortes et al., 2012a).	Functional candidate for GWAS of female traits (Fortes et al., 2012a; Fortes et al., 2013b).
Rhotekin (*RTKN2*)	Family of rhotekin (Collier et al., 2004); codes for rhotekin protein, a part of Rho-binding domain group of Rho-GTPase effectors (Fu et al., 2000), influences exocytosis of the pituitary gland that produces GnRH which regulated the release of LH and FSH (Marques et al., 2000; Momboisse et al., 2011); GWAS study suggests that genetic variation in GnRH stimulation might be influenced by the *RTKN2* gene (Fortes et al., 2012b)	Functional candidate for GWAS of male traits (Fortes et al., 2012b).
Serine peptidase inhibitor, clade A member 7 (*SERPINA7*)	Family of SERPIN (O'Leary et al., 2016) codes for serum thyroid hormone precursor, thyroglobulin (Nonneman et al., 2005) indirectly influences steroid hormone production and spermatogenesis (Nonneman et al., 2005), indirectly affects sex hormone binding globulin level that affects ovarian function (Poppe et al., 2008), and influences the size of testis and sperm count and volume in boars (Dongren et al., 2006; Wagner et al., 2008; Fortes et al., 2012b).	Candidate gene for GWAS in male traits (Fortes et al., 2012b).
SRY-transcription factor 9 (*SOX9*)	Family of the SOX gene (Thomsen et al., 2008) affects male fertility (Thomsen et al., 2008), influences differentiation of sertoli cells in testes, neural crest cells, and chondrocytes (Thomsen et al., 2008), and is associated with testis development. Prostate development and sex determination and sex reversal of embryo (Kent et al., 1996; Vidal et al., 2001; Thomsen et al., 2008; Alankarage et al., 2016).	Candidate gene for GWAS in male traits (Soares et al., 2017).

(Continued on following page)

TABLE 1 | (*Continued*) Genes (identified from literature) as involved in Brahman fertility.

Full Gene Name (Abbreviate gene name)	Function	Evidence for association with fertility in Brahman
Signal transducer and activator of transcription 3 (*STAT3*)	Family of STAT protein involved in the growth hormone receptor (GHR) signaling pathway regulating growth hormone or somatotropin, involved in estrogen receptor pathway, and functional disruption causes infertility, obesity, hyperphagia, and thermal dysregulation (Gao et al., 2004).	Candidate gene for GWAS (Mota et al., 2017) and differential gene expression study (Nguyen et al., 2017) for female traits.
Serine/threonine kinase 11 interacting protein (*STK11IP*)	Affects male fertility (Fortes et al., 2012b), affects spermatogenesis in humans, is associated with inhibin expression in Brahman bulls (Fortes et al., 2012b), and affects spermatozoa development related to histone binding (Steilmann et al., 2011).	Functional candidate for GWAS of male traits (Fortes et al., 2012b).
TATA-box binding protein associated factor 1 (*TAF1*)	Family of TBP-associated factors, TAFs (Freiman et al., 2001), part of TFIID, a basal transcription factor (Freiman et al., 2001), involved in development of normal follicles in the ovary (Freiman et al., 2001), influences regulation of cell cycle and differentiation of spermatids in *Drosophila* (Metcalf and Wassarman, 2007), and affects cattle puberty, time of fertilization, and embryo development (Hecht et al., 2011).	Positional and functional candidate for GWAS of male traits (Fortes et al., 2012a).
TATA-box binding protein associated factor 9b (*TAF9B*)	Family of TBP-associated factors, TAFs (Freiman et al., 2001), functions similar to TAF1 and affects scrotal development in Brahman bulls (Fortes et al., 2012a).	Positional and functional candidate for GWAS of male traits (Fortes et al., 2012a).

$$D = \frac{\left(Ho + \frac{He}{2}\right)}{2,700,000,000}$$

RNA Extraction

Total RNA was isolated using mirVana miRNA Isolation Kit (Ambion) following the manufacturer's instruction. RNA purity was evaluated with a Nanodrop ND-1000 spectrophotometer (v.3.5.2, Thermo Fisher Scientific). QubitTM 4.0 Fluorometer with the Qubit RNA BR (broad-range) assay kit (Thermo Fisher Scientific) was used to quantify RNA concentration. The assessment of RNA integrity was performed using Agilent 2100 Bioanalyser (Agilent Technologies). Only RNA with integrity number greater than 8.0 was used for library preparation for RNA-seq, CAGE-seq, and ISO-seq.

RNA Sequencing

All RNA samples were sent to the Ramaciotti Centre for Genomics (UNSW Sydney, Australia) for library preparation and sequencing. Stranded paired-end RNA-seq libraries were sequenced on a 2 × 100 bp paired-end NovaSeq6000 run with an S4 flowcell. The resulting reads were trimmed in the same way as the whole genome sequencing Illumina data.

CAGE-seq Sequencing

The RNA was sequenced on Illumina single-read flow cells utilizing the 27-nt-long tags prepared corresponding to the 5′-end of the capped RNAs as per Salavati et al., (2020). The libraries were sequenced on an Illumina HiSeq 2500 platform (50 nt single-read) at the Centre for Genomic Research, University of Liverpool, Liverpool. After sequencing, read quality was assessed using FastQC (Andrews, 2010), and quality trimming was administered using Trimmomatic, version 0.35 (Bolger et al., 2014) using the settings "*CROP:9*" to trim the last nine bases and "*HEADCROP:14*" to trim the initial 14 base pairs (Forutan et al., 2021).

ISO-seq Sequencing

First-strand cDNA synthesis was conducted using the TeloPrime Full-Length cDNA Amplification kit (Lexogen, Australia) from 1 μg of total RNA input according to the manufacturer's guideline, with an exemption of the uterus and ovary samples. Due to low extracted RNA concentration, only 500 ng of total RNA from these two tissues was used at this step. Additionally, among these tissue samples, the full-length double-stranded cDNAs from the fetal liver, thyroid, and spleen were prepared using the TeloPrime Full-Length cDNA Amplification version 1 kit (Lexogen, Australia). The TeloPrime Full-Length cDNA Amplification version 2 kit was used for all other samples.

To determine the optimal PCR cycle number for the large-scale PCR, the full-length double-stranded cDNAs were fist amplified in a qPCR reaction using 3′ and 5′ end-specific primers from the TeloPrime Full-Length cDNA Amplification kit (Lexogen, Australia) combined with PCR master mix reagents from PrimerSTAR GXL DNA Polymerase (Takara, Australia). SYBR Green I (Invitrogen, United States) was added to a final concentration of ×0.1 in the qPCR reaction with a total of 40 cycles. QPCR results were evaluated based on the fluorescence value, electrophoresis images, and bioanalyzer results using the Agilent DNA 12000 kit (Agilent Technologies, Germany). Large-scale PCR was then performed using optimal PCR cycles determined during the optimization step for each sample.

PCR products for all samples were sent to Ramaciotti Centre for Genomics (UNSW Sydney, Australia) for library preparation and sequencing. Briefly, PacBio IsoSeq libraries were prepared using the PacBio SMRTBell template prep kit 1.0 SPv3 for sequel protocol. Aliquots of the cDNA products underwent a ×1 and ×0.4 Ampure PB clean-up (Beckman Coulter, Australia). The aliquots were combined post clean-up using different ratios with preference to the aliquots enriched for transcripts above 4 kb. The libraries were sequenced on the PacBio Sequel system using 10-hour movies and v3.0 sequencing chemistry. A total of 11 SMRT cells were sequenced on a PacBio Sequel system (Ramaciotti Centre for Genomics, UNSW Sydney, Australia) for 10 samples. The fetal lung sample was sequenced twice as the first run was overloaded.

Demultiplexing, filtering, and quality control were performed using SMRT Link version 6.0.0 (Pacific Biosciences). The raw reads (subreads) generated by PacBio were used for calling circular consensus sequence (CCS) using the CCS tool (version 4.2.2) with parameters "—skip-polish –min-passess=3." Adapter sequences from these CCS reads were removed using Lima tool (version 1.11.0) with parameters "lima –isoseq –dumclips". The polyA tails and artificial concatemers were trimmed and removed using the refine tool (isoseq refine –require-polya –min-length-polya 8).

Mapping of Expression Data

All three datasets (CAGE-seq, RNA-seq, and ISO-seq) used in this study and the gene sequences were mapped to the Brahman genome that was assembled from the same animal (cow) as all the nonfetal expression data were generated from.

The CAGE-seq data were mapped using BWA-mem (Li and Durbin, 2009) optimized as per Forutan et al. (2021) with the options -M (to mark the shorter split hits as secondary hits), -k 10 (to specify that the sequences with seed length below 10 were skipped), -T 10 (to filter out alignments with a mapping score less than 10), -L 4 (to specify the clipping penalty), and -B 5 (to specify the mismatch penalty).

RNA-seq data were mapped to the Brahman genome using STAR (Dobin et al., 2013). The reference genome was indexed using the "genomeGenerate" option. The alignments were output in sorted BAM format.

Minimap2 (Li, 2018; 2019) was used to align the gene sequences and ISO-seq data. The output was in "SAM" format, and the options "-uf" was used to find canonical splicing sites GT-AG on the transcript strand, "--secondary =no" to skip the output of secondary alignments, and " -C5 -O6,24 -B4" which is the pre-set filter for long-read splice alignment of PacBio circular consensus sequencing reads. The alignments were converted to a sorted bam file using Samtools (Li et al., 2009).

Integrative Genome Viewer (IGV) (Robinson et al., 2011) was used to visualize alignments to interpret transcription start sites, isoforms, and gene expression. The minimum mapping quality was set to 10.

Identifying Heterozygous Loci

The heterozygous loci were identified using whole genome sequencing of the same animals. Single-nucleotide polymorphisms (SNPs) were observed manually in IGV. SNPs were only considered where both alleles were observed in at least two reads.

Statistical Analysis

To calculate the statistical significance of allele-specific expression for each tissue in the RNA-seq data, Pearson's chi-squared test (LaMorte, 2016) was administered by comparing the difference between the expected and observed value with df = 1. The formula is as follows:

For each allele i,

$$\chi^2_{df=1} = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed value and E_i is the expected value for each of the two alleles.

Only loci that were heterozygous, based on the observation of both alleles in the whole genome sequence data, were tested. The ratio of the two alleles at each of the heterozygous loci tested was assumed to be 1:1. Hence, the E-value is calculated by taking the average of the observed value. Therefore,

$$E_i = \frac{R}{2}$$

where E_i is the expected value for the allele i and R is the total number of reads at that locus in the RNA-seq data (sequencing depth).

To control for any sequencing bias between the two alleles, instead of assuming equal allelic ratios, the ratio of alleles observed in the whole genome sequencing data from each of the two animals was also used to determine the expected values. Hence, the expected value for each allele i was

$$E_i = \frac{R \times G_i}{\sum G}$$

where E_i is the expected value for the allele i, R is the total observations at that locus in the RNA-seq data (sequencing depth), G_i is the observations of allele i in the whole genome sequencing (WGS) data, and $\sum G$ is the total number of observations in the WGS data at that locus (sequencing depth).

To compare gene expression between tissues, the number of reads from each dataset (RNA-seq, ISO-seq, and CAGE-seq) was used. The relative expression in reads per million for each gene in each tissue (RPM_{ij}) was calculated as

$$RPM_{ij} = T_{ij} \div A_j \times 1,000,000$$

where T_i is the total number of read pairs mapped to gene i in tissue j and A_j is the total read pairs mapped in tissue j.

RESULTS

Identification of Important Fertility Genes in the Brahman Genome

Thirteen genes important for Brahman fertility were identified in the literature (**Table 1**). The CDS sequence was extracted from

TABLE 2 | List of genes with their respective positions within *Bos taurus* genome and Brahman genome with their alignment length and identity.

Gene	Isoforms	Bos taurus		Bos indicus		Length of gene[a]	Chromosome	Strand direction[b]	Homology (%)
		Start Position	Stop Position	Start Position	Stop Position				
AR	X1	51674157	51881942	84786064	84957236	171172	X	Negative	99.96
IGF	X1	66206081	66263849	66145894	66203733	57839	5	Negative	100
	X2	66206081	66263849	66145894	66203733	57839			99.65
	X3	66206081	66261980	66145894	66199198	53304			100
	X4	66206081	66261980	66145894	66199195	53301			100
	X5	66192424	66263849	66132237	66203733	71496			99.37
	X6	66192595	66263849	66132408	66147706	15298			99.33
	X7	66192424	66261980	66132237	66147706	15469			99.76
	X8	66192424	66261980	66132237	66147706	15469			99.75
	Preprotein	66192424	66263849	66132237	66147706	15469			99.79
INHA	X1	107501844	107504762	107614722	107617642	2920	2	Positive	99.82
PENK	X1	23542677	23546157	23420291	23423774	3483	14	Negative	100
PLAG1	X1	23330541	23332546	23192090	23194095	2005	14	Negative	100
	X2	23330541	23331794	23192090	23193343	1253			
RTKN2	X1	18284769	18393694	17858619	17967072	108453	28	Negative	99.3
	X2	18284769	18385527	17858619	17958855	100236			99.26
	X3	18284769	18348138	17858619	17921964	63345			99.33
SERPINA7	X1	54824445	54829800	53237946	53241468	3522	X	Positive	99.69
	Precursor	54824445	54827949	53237963	53241468	3505			99.68
SOX9	X1	58919579	58922699	60166722	60169843	3121	19	Negative	100
STAT3	X1	42419849	42450618	43645371	43676141	30770	19	Negative	99.66
	X2	42419849	42450618	43645371	43676141	30770			99.66
	X3	42421282	42450618	43647176	43676141	28965			99.64
STK11IP	X1	107521189	107535476	107634090	107648385	14295	2	Positive	99.45
	X2	107521189	107535476	107634090	107648385	11551			99.5
	X3	107521189	107533477	107634090	107646987	12297			99.46
TAF1	X1	79206804	79276760	80853773	80923837	70064	X	Negative	99.86
	X2	79206804	79276760	80853773	80923837	70064			99.86
	X3	79206804	79276760	80853773	80923837	70064			99.79
	X4	79206804	79276760	80853773	80923837	70064			99.84
	X5	79206804	79276760	80853773	80923837	70064			99.84
	X6	79197983	79276760	80857229	80923837	66608			99.87
	X7	79225602	79276760	80872738	80923837	51099			99.9
	X8	79229398	79276760	80885675	80923837	38162			99.9
	X9	79227390	79276760	80885675	80923837	38162			99.9
TAF9B	X1	74232003	74255254	74070058	74093312	23254	X	Positive	100
	X2	74232003	74255924	74070058	74093982	23924			100
	Subunit	74232003	74239263	74070058	74077335	7277			100

[a]*Length from the first base of the first exon to the last base of the last exon.*
[b]*Negative = 3′ to 5′ direction; positive = 5′ to 3′ direction.*

the *Bos taurus* genome and aligned to the Brahman genome assembly using BLASTn to obtain the positions of those genes within the Brahman genome (**Table 2**). The mean homology of the 13 genes between the *Bos taurus* and Brahman genome was 99.76%. The mean homology across the entire genome of the two individuals was 99.50% (**Table 3**). Five genes had at least one isoform that shared a 100% homology between *Bos taurus* and *Brahman* genomes. All isoforms of proenkephalin (*PENK*), pleiomorphic adenoma gene 1 (*PLAG1*), SRY-transcription factor 9 (*SOX9*) and TATA-box binding protein-associated factor 9b (*TAF9B*), and three out of nine isoforms in insulin-like growth factor 1 (*IGF1*) had 100%

TABLE 3 | Genome-wide SNPs compared to *Bos taurus* genome.

	Mother	Foetus
Homozygous SNP[a]	7762707	7734830
Heterozygous SNP	11408849	11537221
Haplotype level Homology	99.501%	99.500%

[a]*Homozygous for alternate alleles from reference.*

TABLE 4 | Position and length of 5′ untranslated region.

Gene	5′ gene position	CAGE-seq TSS peak		UTR length[a]
		Start position	Stop position	
AR	84957236	84958298	84958477	1,241
IGF1	66203733	66203715	66204034	301
INHA	107614722	107614985	107614491	231
PENK	23423774	23424308	23424459	685
RPS20	23140852	23141036	23141219	367
RTKN2	17967072	17966877	17967512	440
SERPINA7	53237946	53236117	53236027	1,919
SOX9	60169843	60169524	60170244	401
STAT3	43676141	43676023	43676320	179
STK11IP	107634090	107634194	107633817	273
TAF1	80923837	80923816	80923981	144
TAF9B	74070058	74070080	74069977	81

[a]*Distance between the start of the coding region within the first exon and the center of the TSS peak from the CAGE-seq data.*

homology. The least homology was observed in the three isoforms of rhotekin (*RTKN2*) ranging from 99.26 to 99.33% (**Table 2**).

Within the 13 genes, heterozygous sites were identified in both the Brahman cow and fetus. There were 270 and 95 SNPs identified within the coding regions of the fertility genes for the cow and fetus, respectively. 55.07% of these were located within the untranslated regions (3.01% in 5′ UTR and 52.05% in the 3′ UTR). The highest number of heterozygous sites was observed in *STK11IP*; no significant heterozygous sites were observed in *TAF1*.

Transcription Start Sites

The transcription start sites (TSSs) were identified using CAGE-seq data (**Supplementary Table S1**) from Forutan et al. (2021), which were remapped to the Brahman genome (Ross et al., 2022). No CAGE-seq peak was identified in the *PLAG1* region. In the 12 out of 13 genes that had a TSS identified, the CAGE-seq peaks were on an average 562 base pairs upstream of the start of the coding region in the first exon in the 5′ direction (**Table 4**). The largest 5′ UTR was found on *SERPINA7,* and the smallest was in *TAF9B*.

Tissue-Specific Expression

Tissue-specific expression of the 13 fertility-related genes was examined by comparing reads per million in all three datasets. In RNA-seq data (**Figure 1**; **Supplementary Table S1**) in blood, *RPS20* has the most expression followed by *STAT3*, *TAF9B,* and *TAF1*. *IGF1*, *SERPINA7,* and *SOX9* were not expressed at detectable levels in the blood sample, and the expression of the rest of the genes was very low. *RPS20* was the most expressed gene in fetal liver followed by *SERPINA7* and the least being *PENK*, *RTKN2,* and *INHA* with rest of the gene expression being insignificant. In adult liver tissue, *STAT3* was the most prominently expressed gene followed by *RPS20*. While other genes showed a relatively similar expression profile between the two liver life stages, the expression level of *AR* was much higher in liver tissue than that in fetal liver. *STAT3* was the most expressed gene in thyroid followed by *RPS20*. The expression of genes *RTKN2*, *STAT3*, *STK11IP*, *TAF1,* and *TAF9B* was higher in adult lung tissue than that in fetal lung, whereas for genes *IGF1*, *INHA*, *PENK*, *PLAG1*, *RPS20,* and *SOX9,* fetal lung showed higher expression. *AR* and *SERPINA7* had no significant expression in both these tissues. *TAF9B* was the most expressed in kidney tissue, while *TAF1* expression is considerably lower. *STAT3* and *RPS20* also had significant expression in kidney tissue. Most genes were lowly expressed in muscle tissue with an exception of *RPS20*, *STAT3*, and *TAF9B*. The spleen and fetal lung showed the highest expression level of *RPS20*, followed by *STAT3* and *TAF9B*, when other genes have considerably lower expression levels. The fetal lung expresses *SOX9* relatively highly, while its expression in the spleen was very low. Overall, fetal lung and spleen had the highest expression across all of the tested genes in the RNA-seq data.

In the CAGE-seq data (**Figure 1**; **Supplementary Table S1**), *RPS20* was the most expressed gene with its highest expression in the fetal lung and ovary. *STK11IP* and *TAF9B* also shows significant expression in all tissues. *PLAG1* had no CAGE-seq expression data, and *STAT3* shows no significant expression in any of the tissues. *SERPINA7* and *INHA* showed significant expression in only the fetal liver and ovary, respectively. *IGF1* was most expressed in the uterus along with significant expression in the liver and lung tissues, ovary, spleen, and thyroid. *AR* showed high expression levels in the ovary and uterus as well as the kidney, liver, and spleen. *SOX9* has a fairly high expression in all tissues except blood and muscle, while *TAF1* is well expressed in all tissues except the adult liver. *PENK* is highly expressed in the uterus in addition with notable expression in the ovary, spleen, and fetal lung.

In the ISO-seq data (**Figure 1**; **Supplementary Tables S1, S2**), *RPS20* was the only gene with detected expression in all tissues, while *PLAG1* showed no expression in any tissue. The ovary and uterus were the tissues where the most genes were expressed despite not having the highest sequencing depth. The ISO-seq data had the lowest sequencing depth of the three technologies used. Given the lower sequencing depth of the ISO-seq data, it is very likely that there are many more uncharacterized isoforms in the data; this is a limitation of this study.

In the ISO-seq data, genes including *PLAG1*, *SERPINA7*, *STAT3*, *SOX9*, and *STK11IP* were highly expressed in the ovary, while *AR*, *IGF1*, *PENK*, *RPS20*, *STAT3*, *SOX9*, and *TAF1*

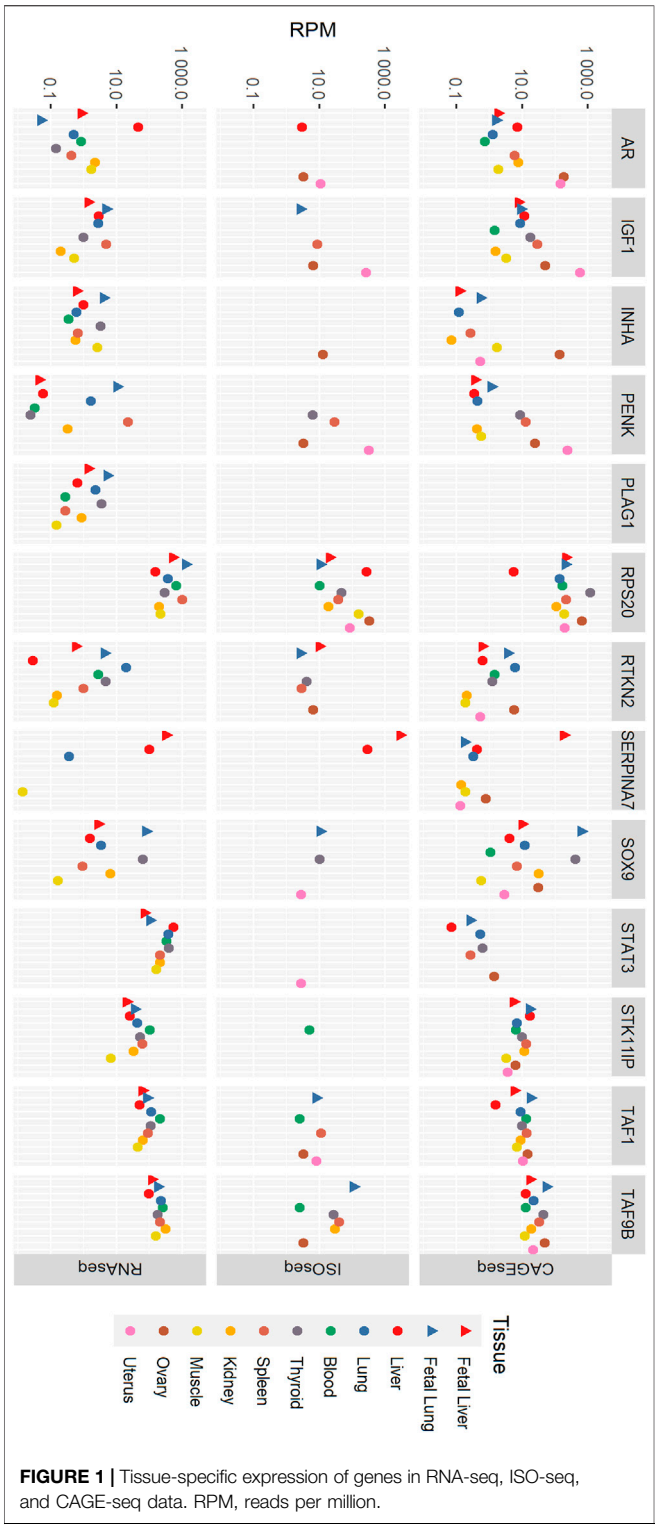


FIGURE 1 | Tissue-specific expression of genes in RNA-seq, ISO-seq, and CAGE-seq data. RPM, reads per million.

were highly expressed in the uterus. Most genes were not expressed in the muscle and kidney. Only *RPS20* was expressed in muscle, and only *RPS20* and *TAF9B* were expressed in the kidney. *TAF9B* was expressed in the blood, fetal lung, kidney, ovary, spleen, and thyroid. *RTKN2* shows expression in the fetal liver but not in the adult liver. It is also expressed in the fetal lung, ovary, spleen, and thyroid. *INHA* is only expressed in the ovary, *STAT3* was only observed in the uterus and *STK11IP* only in blood. Despite *SERPINA7* being expressed only in the liver tissues, it was the most highly expressed gene in the ISO-seq dataset with its highest expression level in the fetal liver. *AR* was expressed in the adult liver, ovary, and uterus and *SOX9* is expressed in the fetal lung, thyroid, and uterus. *IGF1* was expressed in the ovary, spleen, thyroid, and uterus. The highest expression of *PENK* is found in the uterus, while it is also expressed in the fetal lung, ovary, and spleen at lower levels.

Overall, *RPS20* was the most highly expressed gene in any tissue in the RNA-seq, ISO-seq and CAGE-seq data. When the expression level was corrected by length to reads per kilobase million (by dividing by the length of the coding sequence for each gene in kilobasepairs), *RPS20* was still the most highly expressed gene in the RNA-seq data, consistent with result in the CAGE-seq data.

Allele-Specific Expression

Allele-specific expression (**Figure 2**) was tested where there were expression data which overlapped a heterozygous SNP in the relevant animals (the cow or the fetal sample). A total of 117 SNPs were observed in the all genes. *TAF1* had no SNP identified in these two animals. Due to the large number of loci being tested, a p-value cutoff of .0001 was used. Out of the 117 SNPs tested, nine had allele-specific expression in the cow and five in the fetus (chi-squared test; $p < .0001$). To determine the significance of these SNPs was not an effect of the allelic bias in the RNA-seq data; the expected ratios were adjusted to reflect the allelic ratios in the WGS data. In the cow, significant allele-specific expression was still observed in *AR*, *RPS20*, *SERPINA7*, and *TAF9B* (chi-squared test; $p < .01$). The SNP in *SOX9* was not significant when using the new WGS-based ratio. Within the fetal tissues, *RPS20* and *SOX9* showed allele-specific expression when compared to both the 50:50 ratio and WGS-based expected ratios. Of the cow tissues, only the liver displayed allele-specific expression of *AR*, while all tested tissues showed highly significant allele-specific expression in the 5′ UTR of RPS20, which were the fetal tissues. *SERPINA7* showed allele-specific expression in exons 1 and 2 in the liver tissue of the cow, which was the only sample with sufficient coverage to test for ASE. The 3′ UTR of *SOX9* and *TAF9B* had allele-specific expression in both the fetal lung and adult lung. Overall, more SNPs associated with ASE were identified in the 5′ and 3′ UTR of the investigated genes than the exons.

Isoform Discovery

The isoforms present in the genes were observed in the ISO-seq data with supporting evidence from RNA-seq data (**Figure 3**; **Supplementary Tables S3–S6**). For the scope of this study, the isoforms absent in the NCBI database are considered to be novel.

Out of all examined genes, the following had no ISO-seq data mapping to the gene region: *AR*, *PLAG1*, *SOX9*, and *STAT3*. The genes *PENK* and *INHA* had only one isoform each, which were not novel. *STK11IP* had one isoform expressed in blood, but it was not well supported with evidence from RNA-seq.

The highest number of isoforms was observed in *SERPINA7* with seven isoforms expressed in adult and fetal liver tissues. Out

FIGURE 2 | Allele-specific expression for each tissue with significance. Significance at $p < .0001$ and $p < .01$ indicated for reference (dashed lines). The ratio of alleles observed in whole genome sequencing (WGS) of the same animals is indicated as the background of each panel. Only loci that were heterozygous in the WGS data were tested for allele-specific expression. In the position, 1, 2, and 3 refer to exons 1, 2, and 3, respectively.

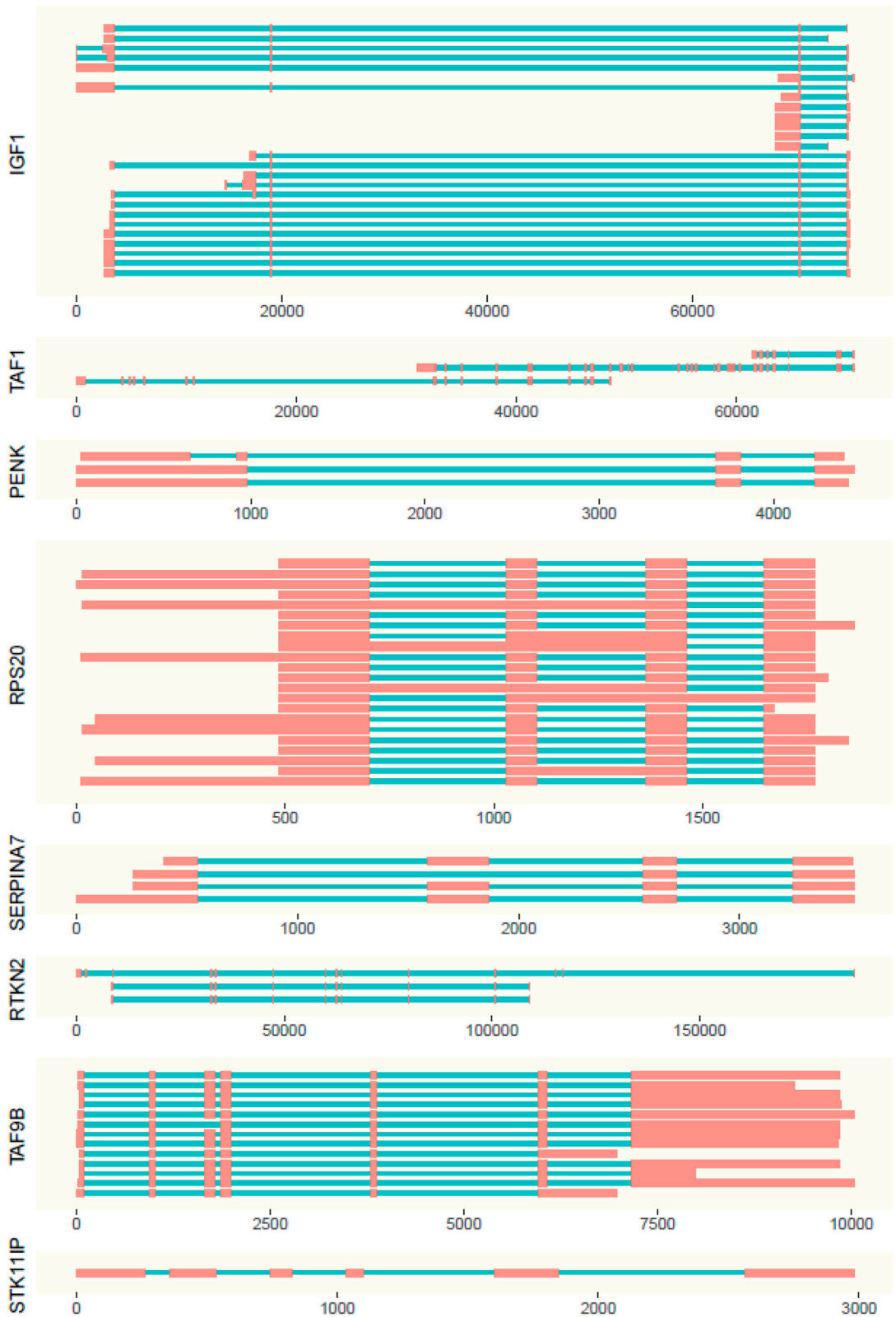


FIGURE 3 | Representation of isoforms with length of exons (red boxes) and introns (blue lines). Positions are displayed relative to the first expressed base pair (x axis) of the gene.

of these isoforms, isoforms I, II, IV, V, VI, and VII were novel (**Figure 3**). Genes *RPS20* and *IGF1* (**Figure 3**) have four isoforms each. Isoform 1 of *RPS20* is present in the fetal liver, kidney, liver, muscle, ovary, spleen, and uterus, while isoform III is present in the kidney and IV in the liver. Isoform II is present in all tissues. Out of these isoforms I, II, and II are found to be novel. In *IGF1*, isoform I is found in the ovary and uterus, while isoform II is there in the spleen and thyroid. Isoforms III and IV, which are both expressed in the uterus, are novel.

The gene *RTKN2* was expressed in the fetal lung, ovary, spleen, and thyroid and has three isoforms out of which all three were novel. Isoform I, with a longer first exon, was found only in the fetal lung and thyroid. Isoform II was observed in the spleen, whereas III was found in the ovary and had the first 2 exons missing (**Figure 3**).

A novel isoform, which missed the first seven exons when compared to the NCBI database of *TAF1* isoforms, was identified for the *TAF1* gene. It is expressed only in the spleen (**Figure 3**).

TAF9B had two isoforms. The absence of the first exon in 5′ direction when compared to the *Bos taurus* gene sequence made both these isoforms novel. Isoform I was present in the blood and ovary and II was observed in the fetal lung, kidney, spleen, and thyroid (**Figure 3**).

Interestingly, *TAF9B* was a partially nested gene. The last exon of *TAF9B* was nested in phosphoglycerate kinase 1 in blood, kidney, lung, muscle, thyroid, fetal liver, and fetal lung tissues.

DISCUSSION

Here, we investigated expression variation of 13 fertility-associated genes in Brahman cattle using RNA-seq, ISO-seq, and CAGE-seq. Within these few genes, the variety of data available allowed us to identify previously unknown levels of variation at the genome and transcriptome level.

The homology between the *Bos taurus* genome annotation and Brahman genome of 99.8% reveals that on average, these genes have a variable site within their expressed regions roughly every 500 bp. This is a slightly higher level of conservation than that of the whole genome, which has 99.5% homology in both the mother and fetus based on identified SNP loci. A higher level of conservation within these genes could be expected given their role in important reproductive traits and the selective pressure in coding or regulatory regions. However, given the extremely small number of animals used in this study, this finding needs to be validated across a large cohort of genetically diverse animals.

Identification of transcription start sites (TSSs) in these genes revealed that (where detected) the TSSs were on an average 562 basepairs upstream of the start of the coding region in the first exon. 5′ UTRs are known to regulate the posttranscriptional modification (Kim et al., 1992; Lawless et al., 2009; Araujo et al., 2012). These play an important role during embryonic development (Van Der Velden and Thomas, 1999; Leppek et al., 2018). 5′ UTRs also regulate the translation of mRNAs (Van Der Velden and Thomas, 1999). Studies suggest that 5′UTR may be utilized to control the expression of genes (Halder et al., 2009). The TSSs within our investigated genes revealed variable 5′ UTR lengths (81–1,919 bp). This is within the observed ranges of 5′ UTR across mammals (Pesole et al., 2001); however, 12 out of the 13 genes had 5′ UTR longer than the reported average for other mammals (nonhuman and nonmouse), while nine out of 13 genes had longer 5′ UTRs than the average reported for humans. Compared to UTR lengths in humans observed from the study by Davuluri et al., (2000), the 5′ UTRs of these genes are substantially longer in most cases (averaging from 47 to 250 bp depending on the expression class). This suggests that these genes related to fertility may have longer than average 5′ UTRs, which may have a role in their regulation (Pesole et al., 2001). Davuluri et al. (2000) found that genes which were thought to be poorly translated tended to have longer 5′ UTRs, suggesting that the translation of these fertility-related genes may be low, especially for *SERINPINA7* and AR, both of which had 5′ UTRs longer than 1KB.

The long 5′ UTR observed in *AR* may lead to low translation. Lyons et al. (2014) suggested that this low level of translation may be overcome by a shift in the transcription start site or alternative splicing. *AR* has a cluster of SNPs and indels in the 5′UTR region in cattle. Within this polymorphism cluster, there are four putative SRY-binding sites in a perfect LD with an SNP associated with scrotal circumference in Brahman bulls (Lyons et al., 2014). Additionally, a knockout study in mice found that *AR* also affects fertility in females (Yeh et al., 2002), suggesting that even though *AR* is primarily involved in male development, it may also contribute to female fertility variation.

This is only the second report of CAGE-seq data analysis in cattle. Here, we used the same dataset as used by Forutan et al. (2021) who reported the structure of TSS positioning within the bovine genome across subspecies and developmental stages. Additionally, a CAGE-seq analysis mapped, identified, and predicted novel and previously unannotated transcription start sites (TSSs) and TSS enhancer cluster (Salavati et al., 2020) in sheep. Here, we took a more specific approach to investigate a subset of genes.

The RNA-seq results revealed that the highest relative expression of these genes is observed in the spleen and fetal lung. This is consistent with the fact that most of these genes considered for this study are related to hormonal regulation and reproductive development (Kent et al., 1996; Le Reith, 1997; Brinkmann et al., 1999; Vidal et al., 2001; Yilmaz et al., 2006; Wang R.-S. et al., 2009). Expression of genes observed in ISO-seq and CAGE-seq data was not as prominent as that in RNA-seq. The sequencing depth within these datasets is much lower than that in RNA-seq, limiting their power to gene expression. It is important to note that a cDNA size selection step was included in the ISO-seq library preparation, which could skew the quantification dramatically. For the CAGE-seq results, the short-read length could limit the ability of the data to map specifically to the genome as there are regions of high homology, such as recent duplications. Therefore, RNA-seq is likely the most accurate representation of relative abundance between the tissues of the three expression datasets used in this study.

The absence of any data for *PLAG1* in ISO-seq and CAGE-seq datasets could be a direct effect of the lack of deep sequencing. Genes such as *AR*, *SERPINA7*, *SOX9*, and *STK11IP* have shown very low expression in all the datasets despite being found to be expressed well in some of these tissues by previous studies (O'Leary et al., 2016). These genes are mainly associated with male fertility, while the available datasets were compiled from a

female specimen (both the cow and the fetus were female); this is the most likely cause of the lower expression levels.

Allele-specific expression was identified in five out of the 13 genes. Liver and fetal liver tissues showed the highest proportion of SNPs with allele-specific expression. Allele-specific expression has been found to be abundant in dairy cattle transcriptome data (Chamberlain et al., 2015), and this analysis confirms that at least for these genes, it is likely the same in Brahman cattle also. Allele-specific expression in a direct mechanism where genetic variation can be linked to phenotypic variation for an important trait as the amount of a gene being expressed can be directly affected by which alleles are present in the animal.

A total of 24 isoforms were found within all the observed genes, out of which 16 are novel. Isoform III of *RPS20* shows an unspliced intron in the ovary and uterus. It is also notable that when intron 3 is retained, intron 2 is always retained, but when intron 2 is retained, intron 3 is not necessarily always retained. Unspliced introns are thought to have a major role in gene expression regulation in plants (Pleiss et al., 2007; Syed et al., 2012), but the part they play in mammals is not fully known. Studies suggest that intron retention may regulate the production of isoforms, stability, and efficiency of translation of RNA and rapid gene expression through posttranscriptional splicing of these introns (Jacob and Smith, 2017). Multiple ISO-seq studies in different species (Xie et al., 2018; Beiki et al., 2019; Feng et al., 2019) have identified that intron retention is quite common as a source of isoform variation.

An important limitation of this study is that only two related animals were used. Given the effect that genetic variation has been observed to have of isoform expression (Garrido-Martín et al., 2021), it could be hypothesized that in a larger population of genetically diverse animals, more variation in isoforms would be observed. The economics of data generation using current technology mean that whole population isoform characterization using long-read sequencing is not currently feasible. However, as sequencing technology continues to improve and costs decrease, population scale isoform discover is likely to be a reality in the near future. Large-scale population datasets will likely show dramatically more variation that has been observed here.

An important aspect of isoform discovery *en masse* is that stringent filters must be placed on the data interpretation to ensure that reported isoforms are robust. This generally means that isoform calling pipelines will collapse down isoforms where the only difference is in the first or last exon or the untranslated regions. The isoform variation in *TAF9B,* where only the length of the untranslated region differs between the two isoforms, is one such example. Usually isoforms such as these would raise suspicion of the shorter isoform being 5′ degraded; however, as there are multiple forms of evidence used to identify these isoforms (ISO-seq and RNA-seq) we can be confident that there are, indeed, two gene forms being expressed. This particular example highlights the usefulness of using multiple data types to identify expression variation.

Despite the absence of the last exon in *TAF9B* in the ISO-seq data, RNA-seq data suggest the presence of an exon downstream in 3′ direction in blood, fetal lung, kidney, ovary, spleen, and thyroid tissues. This is confirmed by a large number of intron spanning RNA-seq reads. This suggests the presence of an exon within another gene (phosphoglycerate kinase 1) present right next to *TAF9B* on the reverse strand. The relevance of this exon on the regulation of expression of both *TAF9B* and the other gene needs to be considered. Studies of nested genes, where genes were fully located within other genes (Yu et al., 2005) showed that the nested genes were under strong selection and displayed reciprocal expression with each other as well as strong tissue-specific expression. We did not observe strong tissue-specific expression to *TAF9B*, and the partially-nested gene it is associated with was also expressed in six of the tissues where *TAF9B* was observed, suggesting that *TAF9B* does not follow the same pattern of expression.

Genes relevant to fertility in Brahman were identified and shown to have tissue-specific expression, allele-specific expression, variation in transcription start sites, untranslated regions, and novel isoforms. This case study is an example of the detailed information that can be obtained from combining information from multiple expression datasets. It is clear that no one datatype is able to fully characterize the transcriptome, and efforts to strategically align data generation efforts will be most beneficial.

DATA AVAILABILITY STATEMENT

CAGE-seq sequencing data are available *via* the European Nucleotide Archive (ENA) under study ID PRJEB44817. The RNA-seq and ISO-seq data for all of the genes presented in the article are available in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

ER, BH, and SM conceived and designed the experiment. LN and YC generated the data. ER and HS analyzed the data and wrote the manuscript. All authors commented on and approved the manuscript.

FUNDING

The authors gratefully acknowledge funding from Meat and Livestock Australia (P.PSH.0868).

ACKNOWLEDGMENTS

This work would not have been possible without support and feedback from the Centre for Animal Science staff and students in the Queensland Alliance for Agriculture and Food Innovation. Special thanks to Dr Bailey Engle for her useful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.784663/full#supplementary-material

REFERENCES

Alankarage, D., Lavery, R., Svingen, T., Kelly, S., Ludbrook, L., Bagheri-Fam, S., et al. (2016). SOX9 Regulates Expression of the Male Fertility Gene Ets Variant Factor 5 (ETV5) during Mammalian Sex Development. *Int. J. Biochem. Cell Biol.* 79, 41–51. doi:10.1016/j.biocel.2016.08.005

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

Araujo, P. R., Yoon, K., Ko, D., Smith, A. D., Qiao, M., Suresh, U., et al. (2012). Before it Gets Started: Regulating Translation at the 5' UTR. *Comp. Funct. Genomics* 2012, 475731. doi:10.1155/2012/475731

Beerda, B., Wyszynska-Koko, J., Te Pas, M. F. W., De Wit, A. A. C., and Veerkamp, R. F. (2008). Expression Profiles of Genes Regulating Dairy Cow Fertility: Recent Findings, Ongoing Activities and Future Possibilities. *Animal* 2, 1158–1167. doi:10.1017/s1751731108002371

Beiki, H., Liu, H., Huang, J., Manchanda, N., Nonneman, D., Smith, T. P. L., et al. (2019). Improved Annotation of the Domestic Pig Genome through Integration of Iso-Seq and RNA-Seq Data. *BMC Genomics* 20, 344. doi:10.1186/s12864-019-5709-y

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., et al. (2018). Meta-analysis of Genome-wide Association Studies for Cattle Stature Identifies Common Genes that Regulate Body Size in Mammals. *Nat. Genet.* 50, 362–367. doi:10.1038/s41588-018-0056-5

Brinkmann, A. O., Blok, L. J., De Ruiter, P. E., Doesburg, P., Steketee, K., Berrevoets, C. A., et al. (1999). Mechanisms of Androgen Receptor Activation and Function. *J. Steroid Biochem. Mol. Biol.* 69, 307–313. doi:10.1016/s0960-0760(99)00049-7

Burns, D. S., Jimenez-Krassel, F., Ireland, J. L. H., Knight, P. G., and Ireland, J. J. (2005). Numbers of Antral Follicles during Follicular Waves in Cattle: Evidence for High Variation Among Animals, Very High Repeatability in Individuals, and an Inverse Association with Serum Follicle-Stimulating Hormone Concentrations1. *Biol. Reprod.* 73, 54–62. doi:10.1095/biolreprod.104.036277

Chamberlain, A. J., Vander Jagt, C. J., Hayes, B. J., Khansefid, M., Marett, L. C., Millen, C. A., et al. (2015). Extensive Variation between Tissues in Allele Specific Expression in an Outbred Mammal. *BMC Genomics* 16, 993. doi:10.1186/s12864-015-2174-0

Collier, F. M., Gregorio-King, C. C., Gough, T. J., Talbot, C. D., Walder, K., and Kirkland, M. A. (2004). Identification and Characterization of a Lymphocytic Rho-GTPase Effector: Rhotekin-2. *Biochem. Biophysical Res. Commun.* 324, 1360–1369. doi:10.1016/j.bbrc.2004.09.205

Davuluri, R. V., Suzuki, Y., Sugano, S., and Zhang, M. Q. (2000). CART Classification of Human 5' UTR Sequences. *Genome Res.* 10, 1807–1816. doi:10.1101/gr.gr-1460r

Dias, M. M., Cánovas, A., Mantilla-Rojas, C., Riley, D. G., Luna-Nevarez, P., Coleman, S. J., et al. (2017). SNP Detection Using RNA-Sequences of Candidate Genes Associated with Puberty in Cattle. *Genet. Mol. Res.* 16. doi:10.4238/gmr16019522

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

Dongren, R., Jun, R., Yuyun, X., Junwu, M., Yanbo, W., Yuanmei, G., et al. (2006). Mutatuion in the Porcine SERPINA7 Gene and its Association with Boar Fertility. *Tpns* 16, 1111–1114. doi:10.1080/10020070612330118

Dowsing, A. T., Yong, E., Clark, M., Mclachlan, R. I., De Kretser, D. M., and Trounson, A. O. (1999). Linkage between Male Infertility and Trinucleotide Repeat Expansion in the Androgen-Receptor Gene. *Lancet* 354, 640–643. doi:10.1016/s0140-6736(98)08413-x

Feng, S., Xu, M., Liu, F., Cui, C., and Zhou, B. (2019). Reconstruction of the Full-Length Transcriptome Atlas Using PacBio Iso-Seq Provides Insight into the Alternative Splicing in Gossypium Australe. *BMC Plant Biol.* 19, 365. doi:10.1186/s12870-019-1968-7

Fortes, M. R. S., Lehnert, S. A., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N. J., et al. (2012a). Finding Genes for Economically Important Traits: Brahman Cattle Puberty. *Anim. Prod. Sci.* 52, 143–150. doi:10.1071/an11165

Fortes, M. R. S., Reverter, A., Hawken, R. J., Bolormaa, S., and Lehnert, S. A. (2012b). Candidate Genes Associated with Testicular Development, Sperm Quality, and Hormone Levels of Inhibin, Luteinizing Hormone, and Insulin-like Growth Factor 1 in Brahman Bulls1. *Biol. Reprod.* 87 (3), 58. doi:10.1095/biolreprod.112.101089

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

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

Forutan, M., Ross, E., Chamberlain, A. J., Nguyen, L., Mason, B., Moore, S., et al. (2021). Evolution of Tissue and Developmental Specificity of Transcription Start Sites in Bos taurus indicus. *Commun. Biol.* 4, 829.

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E., and Tjian, R. (2001). Requirement of Tissue-Selective TBP-Associated Factor TAF II 105 in Ovarian Development. *Science* 293, 2084–2087. doi:10.1126/science.1061935

Fu, Q., Yu, L., Liu, Q., Zhang, J., Zhang, H., and Zhao, S. (2000). Molecular Cloning, Expression Characterization, and Mapping of a Novel Putative Inhibitor of Rho GTPase Activity, RTKN, to D2S145-D2s286. *Genomics* 66, 328–332. doi:10.1006/geno.2000.6212

Gao, Q., Wolfgang, M. J., Neschen, S., Morino, K., Horvath, T. L., Shulman, G. I., et al. (2004). Disruption of Neural Signal Transducer and Activator of Transcription 3 Causes Obesity, Diabetes, Infertility, and thermal Dysregulation. *Proc. Natl. Acad. Sci.* 101, 4661–4666. doi:10.1073/pnas.0303992101

Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F., and Guigó, R. (2021). Identification and Analysis of Splicing Quantitative Trait Loci across Multiple Tissues in the Human Genome. *Nat. Commun.* 12, 727. doi:10.1038/s41467-020-20578-2

Gonzalez-Garay, M. L. (2016). *Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq)*. Editor J. Wu (Dordrecht: Springer Netherlands), 141–160. doi:10.1007/978-94-017-7450-5_6:

Halder, K., Wieland, M., and Hartig, J. S. (2009). Predictable Suppression of Gene Expression by 5′-UTR-Based RNA Quadruplexes. *Nucleic Acids Res.* 37, 6811–6817. doi:10.1093/nar/gkp696

Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annu. Rev. Anim. Biosci.* 7, 89–102. doi:10.1146/annurev-animal-020518-115024

Hecht, N., Cavalcanti, M. C. O., Nayudu, P., Behr, R., Reichenbach, M., Weidner, W., et al. (2011). Protamine-1 Represents a Sperm Specific Gene Transcript: a Study in *Callithrix jacchus* and *Bos taurus*. *Andrologia* 43, 167–173. doi:10.1111/j.1439-0272.2009.01038.x

Jacob, A. G., and Smith, C. W. J. (2017). Intron Retention as a Component of Regulated Gene Expression Programs. *Hum. Genet.* 136, 1043–1057. doi:10.1007/s00439-017-1791-x

Juma, A. R., Damdimopoulou, P. E., Grommen, S. V. H., Van De Ven, W. J. M., and De Groef, B. (2016). Emerging Role of PLAG1 as a Regulator of Growth and Reproduction. *J. Endocrinol.* 228, R45–R56. doi:10.1530/joe-15-0449

Kaneko, H. (2016). "Inhibin," in *Handbook of Hormones*. Editors Y. Takei, H. Ando, and K. Tsutsui (San Diego: Academic Press), 292–294. doi:10.1016/b978-0-12-801028-0.00187-2

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A. C., Baurain, D., et al. (2011). Variants Modulating the Expression of a Chromosome Domain Encompassing PLAG1 Influence Bovine Stature. *Nat. Genet.* 43, 405–413. doi:10.1038/ng.814

Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H., and Koopman, P. (1996). A Male-specific Role for SOX9 in Vertebrate Sex Determination. *Development* 122, 2813–2822. doi:10.1242/dev.122.9.2813

Kim, S. J., Park, K., Koeller, D., Kim, K. Y., Wakefield, L. M., Sporn, M. B., et al. (1992). Post-transcriptional Regulation of the Human Transforming Growth Factor-Beta 1 Gene. *J. Biol. Chem.* 267, 13702–13707. doi:10.1016/s0021-9258(18)42270-3

Lamorte, W. W. (2016). Hypothesis Testing - Chi Squared Test [Online]. Available at: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ChiSquare/BS704_HypothesisTesting-ChiSquare3.html.

Lawless, C., Pearson, R. D., Selley, J. N., Smirnova, J. B., Grant, C. M., Ashe, M. P., et al. (2009). Upstream Sequence Elements Direct post-transcriptional Regulation of Gene Expression Under Stress Conditions in Yeast. *BMC Genom.* 10, 7. doi:10.1186/1471-2164-10-7

Le Reith, D. (1997). Seminars in Medicine of the Beth Israel Deaconess Medical Center. Insulin-like Growth Factors. *N. Engl. J. Med.* 336, 633–640. doi:10.1056/NEJM199702273360907

Leppek, K., Das, R., and Barna, M. (2018). Functional 5′ UTR mRNA Structures in Eukaryotic Translation Regulation and How to Find Them. *Nat. Rev. Mol. Cell Biol.* 19, 158–174. doi:10.1038/nrm.2017.103

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

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

Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191

Littlejohn, M., Grala, T., Sanders, K., Walker, C., Waghorn, G., Macdonald, K., et al. (2012). Genetic Variation in PLAG1 Associates with Early Life Body Weight and Peripubertal Weight and Growth in *Bos taurus*. *Anim. Genet.* 43, 591–594. doi:10.1111/j.1365-2052.2011.02293.x

Lyons, R. E., Loan, N. T., Dierens, L., Fortes, M. R. S., Kelly, M., Mcwilliam, S. S., et al. (2014). Evidence for Positive Selection of Taurine Genes within a QTL Region on Chromosome X Associated with Testicular Size in Australian Brahman Cattle. *BMC Genet.* 15, 6. doi:10.1186/1471-2156-15-6

Maclean, H. E., Warne, G. L., and Zajac, J. D. (1995). Defects of Androgen Receptor Function: from Sex Reversal to Motor Neurone Disease. *Mol. Cell Endocrinol.* 112, 133–141. doi:10.1016/0303-7207(95)03608-a

Malven, P. (1995). Role of Endogenous Opioids for Regulation of the Oestrous Cycle in Sheep and Cattle. *Reprod. Domest. Anim.* 30, 183–187. doi:10.1111/j.1439-0531.1995.tb00143.x

Marques, P., Skorupskaite, K., Rozario, K. S., Anderson, R. A., and George, J. T. (2000). "Physiology of GnRH and Gonadotropin Secretion," in *Endotext*. Editors K. R. Feingold, B. Anawalt, A. Boyce, G. Chrousos, W. W. de Herder, K. Dhatariya, et al. (South Dartmouth (MA): MDText.com, Inc). Available at: https://www.ncbi.nlm.nih.gov/books/NBK279070/.

Mcgowan, K. A., Li, J. Z., Park, C. Y., Beaudry, V., Tabor, H. K., Sabnis, A. J., et al. (2008). Ribosomal Mutations Cause P53-Mediated Dark Skin and Pleiotropic Effects. *Nat. Genet.* 40, 963–970. doi:10.1038/ng.188

Metcalf, C. E., and Wassarman, D. A. (2007). Nucleolar Colocalization of TAF1 and Testis-specific TAFs duringDrosophilaspermatogenesis. *Dev. Dyn.* 236, 2836–2843. doi:10.1002/dvdy.21294

Minten, M. A., Bilby, T. R., Bruno, R. G. S., Allen, C. C., Madsen, C. A., Wang, Z., et al. (2013). Effects of Fertility on Gene Expression and Function of the Bovine Endometrium. *PLoS One* 8, e69444. doi:10.1371/journal.pone.0069444

Momboisse, F., Houy, S., Ory, S., Calco, V., Bader, M. F., and Gasman, S. (2011). How Important Are Rho GTPases in Neurosecretion? *J. Neurochem.* 117, 623–631. doi:10.1111/j.1471-4159.2011.07241.x

Moore, S. G., Pryce, J. E., Hayes, B. J., Chamberlain, A. J., Kemper, K. E., Berry, D. P., et al. (2016). Differentially Expressed Genes in Endometrium and Corpus Luteum of Holstein Cows Selected for High and Low Fertility Are Enriched for Sequence Variants Associated with Fertility1. *Biol. Reprod.* 94. doi:10.1095/biolreprod.115.132951

Mota, R. R., Guimarães, S. E. F., Fortes, M. R. S., Hayes, B., Silva, F. F., Verardo, L. L., et al. (2017). Genome-wide Association Study and Annotating Candidate Gene Networks Affecting Age at First Calving in Nellore Cattle. *J. Anim. Breed. Genet.* 134, 484–492. doi:10.1111/jbg.12299

Müller, M.-P., Rothammer, S., Seichter, D., Russ, I., Hinrichs, D., Tetens, J., et al. (2017). Genome-wide Mapping of 10 Calving and Fertility Traits in Holstein Dairy Cattle with Special Regard to Chromosome 18. *J. Dairy Sci.* 100, 1987–2006. doi:10.3168/jds.2016-11506

National Center for Biotechnology Information (2008). BLAST® Command Line Applications User Manual [Online].

Nguyen, L., Reverter, A., Cánovas, A., Porto-Neto, L., Venus, B., Islas-Trejo, A., et al. (2019). Pre-and Post-Puberty Co-expression Gene Networks in Brahman Heifers Using RNA-Sequencing.

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

Nguyen, L. T., Reverter, A., Cánovas, A., Venus, B., Islas-Trejo, A., Porto-Neto, L. R., et al. (2017). Global Differential Gene Expression in the Pituitary Gland and the Ovaries of Pre- and Postpubertal Brahman Heifers1. *J. Anim. Sci.* 95, 599–615. doi:10.2527/jas.2016.0921

Nonneman, D., Rohrer, G. A., Wise, T. H., Lunstra, D. D., and Ford, J. J. (2005). A Variant of Porcine Thyroxine-Binding Globulin Has Reduced Affinity for Thyroxine and Is Associated with Testis Size1. *Biol. Reprod.* 72, 214–220. doi:10.1095/biolreprod.104.031922

O'leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., Mcveigh, R., et al. (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* 44, D733–D745. doi:10.1093/nar/gkv1189

Ozsolak, F., and Milos, P. M. (2011). RNA Sequencing: Advances, Challenges and Opportunities. *Nat. Rev. Genet.* 12, 87–98. doi:10.1038/nrg2934

Pausch, H., Flisikowski, K., Jung, S., Emmerling, R., Edel, C., Götz, K.-U., et al. (2011). Genome-wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related Traits in Cattle. *Genetics* 187, 289–297. doi:10.1534/genetics.110.124057

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., and Liuni, S. (2001). Structural and Functional Features of Eukaryotic mRNA Untranslated Regions. *Gene* 276, 73–81. doi:10.1016/s0378-1119(01)00674-6

Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. (2007). Rapid, Transcript-specific Changes in Splicing in Response to Environmental Stress. *Mol. Cell* 27, 928–937. doi:10.1016/j.molcel.2007.07.018

Poppe, K., Velkeniers, B., and Glinoer, D. (2008). The Role of Thyroid Autoimmunity in Fertility and Pregnancy. *Nat. Rev. Endocrinol.* 4, 394–405. doi:10.1038/ncpendmet0846

Richards, J. S., and Pangas, S. A. (2010). The Ovary: Basic Biology and Clinical Implications. *J. Clin. Invest.* 120, 963–972. doi:10.1172/jci41350

Robinson, A. J., and Ross, E. M. (2019). QuAdTrim: Overcoming Computational Bottlenecks in Sequence Quality Control. *bioRxiv*. doi:10.1101/2019.12.18.870642

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., et al. (2020). De Novo assembly of the Cattle Reference Genome with Single-Molecule Sequencing. *GigaScience* 9, giaa021. doi:10.1093/gigascience/giaa021

Ross, E. M., Nguyen, L. T., Lamb, H. J., Moore, S. S., and Hayes, B. J. (2022). The Genome of Tropically Adapted Brahman Cattle (*Bos taurus indicus*) Reveals Novel Genome Variation in Production Animals. *bioRxiv*. doi:10.1101/2022.02.09.479458

Salavati, M., Caulton, A., Clark, R., Gazova, I., Smith, T. P. L., Worley, K. C., et al. (2020). Global Analysis of Transcription Start Sites in the New Ovine Reference Genome (Oar Rambouillet v1.0). *Front. Genet.* 11. doi:10.3389/fgene.2020.580580

Simmen, R. C. M., Ko, Y., and Simmen, F. A. (1993). Insulin-like Growth Factors and Blastocyst Development. *Theriogenology* 39, 163–175. doi:10.1016/0093-691x(93)90031-y

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

Soares, A. C. C., Guimarães, S. E. F., Kelly, M. J., Fortes, M. R. S., e Silva, F. F., Verardo, L. L., et al. (2017). Multiple-trait Genomewide Mapping and Gene Network Analysis for Scrotal Circumference Growth Curves in Brahman Cattle. *J. Anim. Sci.* 95, 3331–3345. doi:10.2527/jas2017.1409

Steilmann, C., Paradowska, A., Bartkuhn, M., Vieweg, M., Schuppe, H.-C., Bergmann, M., et al. (2011). Presence of Histone H3 Acetylated at Lysine 9 in Male Germ Cells and its Distribution Pattern in the Genome of

Human Spermatozoa. *Reprod. Fertil. Dev.* 23, 997–1011. doi:10.1071/rd10197

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* 54, 1–33.30.31-31.30.33. doi:10.1002/cpbi.5

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative Splicing in Plants - Coming of Age. *Trends Plant Sci.* 17, 616–623. doi:10.1016/j.tplants.2012.06.001

Takahashi, A. (2016). "Enkephalin," in *Handbook of Hormones*. Editors Y. Takei, H. Ando, and K. Tsutsui (San Diego, CA: Academic Press), 55–e7AA-52. doi:10.1016/b978-0-12-801028-0.00117-3*:*

Takahashi, H., Kato, S., Murata, M., and Carninci, P. (2012a). CAGE (Cap Analysis of Gene Expression): a Protocol for the Detection of Promoter and Transcriptional Networks. *Methods Mol. Biol.* 786, 181–200. doi:10.1007/978-1-61779-292-2_11

Takahashi, H., Lassmann, T., Murata, M., and Carninci, P. (2012b). 5′ End-Centered Expression Profiling Using Cap-Analysis Gene Expression and Next-Generation Sequencing. *Nat. Protoc.* 7, 542–561. doi:10.1038/nprot.2012.005

Taylor, J. A., Goubillon, M.-L., Broad, K. D., and Robinson, J. E. (2007). Steroid Control of Gonadotropin-Releasing Hormone Secretion: Associated Changes in Pro-opiomelanocortin and Preproenkephalin Messenger RNA Expression in the Ovine Hypothalamus1. *Biol. Reprod.* 76, 524–531. doi:10.1095/biolreprod.106.055533

Thomsen, M. K., Francis, J. C., and Swain, A. (2008). The Role of Sox9 in Prostate Development. *Differentiation* 76, 728–735. doi:10.1111/j.1432-0436.2008.00293.x

Ulloa-Aguirre, A., Reiter, E., and Crépieux, P. (2018). FSH Receptor Signaling: Complexity of Interactions and Signal Diversity. *Endocrinology* 159, 3020–3035. doi:10.1210/en.2018-00452

Utsunomiya, Y. T., Milanesi, M., Utsunomiya, A. T. H., Torrecilha, R. B. P., Kim, E.-S., Costa, M. S., et al. (2017). A PLAG1 Mutation Contributed to Stature Recovery in Modern Cattle. *Sci. Rep.* 7, 17140. doi:10.1038/s41598-017-17127-1

Van Der Velden, A. W., and Thomas, A. A. M. (1999). The Role of the 5′ Untranslated Region of an mRNA in Translation Regulation during Development. *Int. J. Biochem. Cell Biol.* 31, 87–106. doi:10.1016/s1357-2725(98)00134-4

Velazquez, M. A., Spicer, L. J., and Wathes, D. C. (2008). The Role of Endocrine Insulin-like Growth Factor-I (IGF-I) in Female Bovine Reproduction. *Domest. Anim. Endocrinol.* 35, 325–342. doi:10.1016/j.domaniend.2008.07.002

Vidal, V. P. I., Chaboissier, M.-C., De Rooij, D. G., and Schedl, A. (2001). Sox9 Induces Testis Development in XX Transgenic Mice. *Nat. Genet.* 28, 216–217. doi:10.1038/90046

Wagner, M. S., Wajner, S. M., and Maia, A. L. (2008). The Role of Thyroid Hormone in Testicular Development and Function. *J. Endocrinol.* 199, 351–365. doi:10.1677/joe-08-0218

Wang, R.-S., Yeh, S., Tzeng, C.-R., and Chang, C. (2009). Androgen Receptor Roles in Spermatogenesis and Fertility: Lessons from Testicular Cell-specific Androgen Receptor Knockout Mice. *Endocr. Rev.* 30, 119–132. doi:10.1210/er.2008-0025

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: a Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484

Wilson, M. (1998). Premature Elevation in Serum Insulin-like Growth Factor-I Advances First Ovulation in Rhesus Monkeys. *J. Endocrinol.* 158, 247–257. doi:10.1677/joe.0.1580247

Xie, S. Q., Han, Y., Chen, X. Z., Cao, T. Y., Ji, K. K., Zhu, J., et al. (2018). ISOdb: A Comprehensive Database of Full-Length Isoforms Generated by Iso-Seq. *Int. J. Genomics* 2018, 9207637. doi:10.1155/2018/9207637

Yeh, S., Tsai, M.-Y., Xu, Q., Mu, X.-M., Lardy, H., Huang, K.-E., et al. (2002). Generation and Characterization of Androgen Receptor Knockout (ARKO) Mice: An *In Vivo* Model for the Study of Androgen Functions in Selective Tissues. *Proc. Natl. Acad. Sci.* 99, 13498–13503. doi:10.1073/pnas.212474399

Yilmaz, A., Davis, M. E., and Simmen, R. C. M. (2006). Analysis of Female Reproductive Traits in Angus Beef Cattle Divergently Selected for Blood Serum Insulin-Like Growth Factor I Concentration. *Theriogenology* 65, 1180–1190. doi:10.1016/j.theriogenology.2005.06.018

Yilmaz, A., Davis, M. E., and Simmen, R. C. M. (2004). Estimation of (Co)variance Components for Reproductive Traits in Angus Beef Cattle Divergently Selected for Blood Serum IGF-I Concentration123. *J. Anim. Sci.* 82, 2285–2292. doi:10.2527/2004.8282285x

Yu, P., Ma, D., and Xu, M. (2005). Nested Genes in the Human Genome. *Genomics* 86, 414–422. doi:10.1016/j.ygeno.2005.06.008

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* 7, 203–214. doi:10.1089/10665270050081478

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 in Genetics

ORIGINAL RESEARCH
published: 28 March 2022
doi: 10.3389/fgene.2022.828884

Genome-Wide Association Studies and Haplotype-Sharing Analysis Targeting the Egg Production Traits in Shaoxing Duck

Wenwu Xu[1], Zhenzhen Wang[1], Yuanqi Qu[2], Qingyi Li[2], Yong Tian[1], Li Chen[1], Jianhong Tang[2], Chengfeng Li[2], Guoqin Li[1], Junda Shen[1], Zhengrong Tao[1], Yongqing Cao[1], Tao Zeng[1] and Lizhi Lu[1]**

[1]Institute of Animal Husbandry and Veterinary Science, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]Hubei Shendan Co., Ltd., Wuhan, China

OPEN ACCESS

Edited by:
Aixin Liang,
Huazhong Agricultural University, China

Reviewed by:
George R. Wiggans,
Council on Dairy Cattle Breeding, United States
Lei Zhou,
China Agricultural University, China

***Correspondence:**
Tao Zeng
zengtao4009@126.com
Lizhi Lu
lulizhibox@163.com

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

Received: *04 December 2021*
Accepted: *11 February 2022*
Published: *28 March 2022*

Citation:
Xu W, Wang Z, Qu Y, Li Q, Tian Y, Chen L, Tang J, Li C, Li G, Shen J, Tao Z, Cao Y, Zeng T and Lu L (2022) Genome-Wide Association Studies and Haplotype-Sharing Analysis Targeting the Egg Production Traits in Shaoxing Duck. Front. Genet. 13:828884. doi: 10.3389/fgene.2022.828884

Age at first egg (AFE) and egg number (EN) are economically important traits related to egg production, as they directly influence the benefits of the poultry industry, but the molecular genetic research that affects those traits in laying ducks is still sparse. Our objective was to identify the genomic regions and candidate genes associated with AFE, egg production at 43 weeks (EP43w), and egg production at 66 weeks (EP66w) in a Shaoxing duck population using genome-wide association studies (GWASs) and haplotype-sharing analysis. Single-nucleotide polymorphism (SNP)-based genetic parameter estimates showed that the heritability was 0.15, 0.20, and 0.22 for AFE, EP43w, and EP66w, respectively. Subsequently, three univariate GWASs for AFE, EP43w, and EP66w were carried out independently. Twenty-four SNPs located on chromosome 25 within a 0.01-Mb region that spans from 4.511 to 4.521 Mb were associated with AFE. There are two CIs that affect EP43w, i.e., twenty-five SNPs were in strong linkage disequilibrium region spanning from 3.186 to 3.247 Mb on chromosome 25, a region spanning from 4.442 to 4.446 Mb on chromosome 25, and two interesting genes, ACAD8 and THYN1, that may affect EP43w in laying ducks. There are also two CIs that affect EP66w, i.e., a 2.412-Mb region that spans from 127.497 to 129.910 Mb on chromosome 2 and a 0.355-Mb region that spans from 4.481 to 4.837 Mb on chromosome 29, and CA2 and GAMT may be the putative candidate genes. Our study also found some haplotypes significantly associated with these three traits based on haplotype-sharing analysis. Overall, this study was the first publication of GWAS on egg production in laying ducks, and our findings will be helpful to provide some candidate genes and haplotypes to improve egg production performance based on breeding in laying duck. Additionally, we learned from a method called bootstrap test to verify the reliability of a GWAS with small experimental samples that users can access at https://github.com/xuwenwu24/Bootstrap-test.

Keywords: genome-wide association study, haplotype, laying duck, egg production, gene

INTRODUCTION

Egg production traits, including age at first egg (AFE) and egg number (EN), have always been a focus of attention in laying ducks, as they directly affect economic benefits to farmers. EN has experienced considerable genetic progress in commercial laying ducks breeds through traditional selection for several decades, reaching a level at an egg on almost every day in highly efficient laying ducks. AFE is also a very important trait for egg production, as it is a partial determination of the laying period. Nowadays, young laying ducks as early as 16 weeks of age start to produce their first egg.

So far, egg production has been greatly improved through the conventional selection strategy. However, the conventional breeding approaches are greatly influenced by the environmental effects, which unavoidably lead to inaccurate heritability estimation (Meuwissen et al., 2001; Muir 2007). We can dissect and quantify the genetic variations in egg production traits with the development of high-throughput genotyping platforms, and the genetic gain in egg production traits can be greatly increased by using a new molecular breeding strategy. Thus, identifying genetic variants affecting egg production traits is one of the primary goals in duck genetics. With the advances in technologies of molecular genetics and availability of single-nucleotide polymorphism (SNP) markers, numerous studies had been conducted to identify quantitative trait loci (QTLs) and SNPs that are associated with EN in poultry. The AnimalQTLdb website (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index) reported 185 QTLs on 24 different chromosomes associated with AFE, EN, and egg production rate in chickens (Tuiskula-Haavisto et al., 2002; Sasaki et al., 2004; Schreiweis et al., 2006; Atzmon et al., 2008; Goto et al., 2011; Xu et al., 2011; Goraga et al., 2012). However, the molecular genetic research that affects egg-laying performance in laying ducks is still sparse, with only a few candidate gene studies. Some researchers have found some candidate genes, such as OIH, FSHβ, GnIH, FSHR, LRP8, VLDLR, and HSP90, to be associated with egg-laying performance in laying ducks (Xu et al., 2011; Wang et al., 2013). There are some limitations in the candidate gene study, such as the uncertainty of candidate gene selection, different genes could be heterogeneous in populations with different genetic backgrounds, the number of gene annotations of duck is small, and the function of some annotated genes is still not completely known today (Kwon and Goate 2000). Therefore, candidate gene study could not be fully utilized to analyze the molecular genetic mechanism of egg production in laying ducks.

Genome-wide association study (GWAS) has become an exceedingly effective and widely used approach in the identification of genetic variants associated with complex traits since the first application of GWAS research on age-related macular degeneration was performed successfully in 2005 by Klein et al. (Klein et al., 2005). Shaoxing duck is an excellent and high-yielding egg breed of duck, and the feeding rate reached 60% in China. After breeding, the age at the first egg of Shaoxing ducks is about 130 days, and the annual egg production can reach 300. In this study, we employed 10× whole-genome sequencing to identify the genomic regions and candidate genes associated with AFE, egg production at 43 weeks (EP43w), and egg production at 66 weeks (EP66w) in a pure line population derived from Shaoxing duck using GWASs and haplotype-sharing analyses, which could potentially accelerate the genetic improvement of egg production.

MATERIALS AND METHODS

Ducks and Phenotypes

A total number of 166 Shaoxing ducks from Hubei Shendan Co., Ltd. (Wuhan, China) were used in our study. Blood samples were collected from brachial veins using the standard procedure in week 66. All ducks were housed in individual cages of the same condition. The AFE and weekly egg production from the onset of laying eggs to 66 weeks of age for each duck were recorded, and then the data were used to define two egg production traits, as the EN from the onset of laying eggs to 43 weeks (EP43w) and the EN from the onset of laying eggs to 66 weeks (EP66w). Animal care and use protocol was approved by the Institutional Animal Care and Use Committee of the Zhejiang Academy of Agricultural Sciences (approval number: 2021ZAASLA15), which was in accordance with the Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China).

Genome Sequencing

A standard cetyl trimethylammonium bromide (CTAB) method was used to isolate genomic DNA from blood, and agarose gel electrophoresis was used to examine the quality and quantity of DNA. After the examinations, paired-end libraries were generated for each eligible sample using standard procedures. Fragments were end-repaired, A-tailed, ligated to paired-end adaptors, and PCR amplified with 500-bp inserts for library construction. According to the manufacturer's standard protocols, libraries were subjected to 150-bp paired-end sequencing on a HiSeq platform (Illumina, San Diego, CA, USA), to a mean sequencing depth of 10× for experimental animals. The depth ensured the accuracy of variant calling and genotyping and met the requirements for population genetic analyses.

Variant Discovery and Genotyping

The 150-bp paired-end raw reads were aligned to the reference duck genome assembly CAU_duck1.0 with the Burrows–Wheeler alignment (BWA aln) using default parameters (Li and Durbin 2009; Huang et al., 2013). On average, 96.4% of the reads were mapped, resulting in a final average sequencing coverage of ×10 (ranging from ×8 to ×18) per individual. Mapping details of 166 resequencing samples were shown in **Supplementary File S1**: **Supplementary Tables S1–S3**. The paired reads that were mapped to the exact same position on the reference genome were marked and removed by Picard MarkDuplicates (http://broadinstitute.github.io/picard) to avoid any influence on variant detection. For GATK SNP calling, standard preprocessing (including realignment and recalibration) and calling

procedures were used (DePristo et al., 2011), each sample generated its own gVCF file, and the files were merged. The output file was further filtered using VCFtools with the filter expression as QUAL < 30, QD < 2.0, MQ < 40, and FS > 60 (Danecek et al., 2011). SNPs that did not meet the following criteria were excluded: 1) a minor allele frequency >0.05; 2) maximum missing rate <0.1; and 3) only two genotypes. Identified SNPs were further classified by SnpEff based on the gene annotation of the reference genome (Cingolani et al., 2012).

Single-Trait Genome-Wide Association Study Analysis

GEMMA (v.0.94) was employed for the single-marker association test between variants and phenotypes underlining a univariate linear mixed model (see **Eq. 1**) and is described in the following equation (Zhou and Stephens 2012):

$$y = W\alpha + x\beta + u + \epsilon; u \sim MVN_n(0, \lambda\tau^{-1}K),\ \epsilon \sim MVN_n(0, \lambda\tau^{-1}I_n) \quad (1)$$

$$FDR(P_i) = P_i * \frac{m}{K_{(P_i)}} \quad (2)$$

where y is the vector of phenotypic observation (AFE, EP43w, and EP66w); W is a design matrix of fixed effect, including a column of 1 s; α is a vector of fixed effects; x is a matrix of genotypes; β is the effect of SNPs; u is a vector of random effects following the multivariate normal distribution $MVN_n(0, \lambda\tau^{-1}K)$, in which λ is the ratio is between τ^{-1} and the variance of polygenetic effects, τ^{-1} is the variance of the residual errors, and K is a kinship matrix estimated from whole-genome sequence variants; ϵ is a vector of errors following the multivariate normal distribution (see **Eq. 1**), and I_n is an identity matrix. With high-density markers throughout the whole genome, naïve Bonferroni corrections of 0.05 divided by the number of examined SNPs to correct multiple comparisons would lead to an overly conservative threshold in our study due to the SNPs being highly correlated with each other. The empirical distribution of *p*-values of markers was used to calculate the genome-wide false discovery rate (FDR) following Storey and Benjamini (Efron and Tibshirani 1993; Yoav and Daniel 2001). The mathematic expression of the FDR is shown in **Eq. 2**, where m is the number of markers, P_i is the *p*-value of the *i*th marker, and $K_{(P_i)}$ is the *p*-value of the *i*th marker ranked in all markers. Population stratification is one of the factors that affect the validity of a GWAS (Pearson and Manolio 2008). Quantile–quantile plots (Q-Q plots) were implemented to evaluate population stratification effects and were constructed with R software to check if stratification exists in our results.

Post Genome-Wide Association Analysis

To detect the linkage disequilibrium (LD) of SNPs near the most significant SNPs in the GWAS results, the 3-Mb region near the top SNPs in the whole-sequence association results was used to conduct LD analysis by extracting genotypes from the data set using plink 1.07 (Pearson and Manolio 2008), and the default settings for minimum linkage between SNPs were at threshold r^2 = 0.4. After the CIs were determined, an investigation of gene ontology (GO) for the genes within the CI was performed to determine biological processes associated with traits using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/home.jsp) (Huang da et al., 2009).

In addition, the haplotypes in the CI were constructed by fastphase with the default setting, and an attempt to find the sharing susceptibility haplotype was made by thoroughly scanning the haplotypes of all individuals (Scheet and Stephens 2006).

TABLE 1 | Descriptive statistic for phenotype values.

Traits	N	Min	Max	Mean	SD
AFE	166	105.00	172.00	136.95	13.93
EP43w	166	113.00	189.00	151.27	12.59
EP66w	166	194.00	325.00	268.96	24.37

Note. AFE, age at first egg; EP43w, the egg number from onset of laying eggs to 43 weeks; EP66w, the egg number from onset of laying eggs to 66 weeks; N, number of samples; Min, the minimum of phenotype values; Max, the maximum of phenotype values.

Bootstrap Test

In this study, the bootstrap test was carried out to verify the reliability of GWASs, which was a resampling technique used to estimate statistics on a population by sampling a dataset with replacement. This method can be used to estimate summary statistics such as the mean, SD, CI, or correlation coefficient, which is done by repeatedly taking small samples, calculating the statistic, and taking the average of the calculated statistics. There were two steps for the bootstrap test in this study; first, random resampling was performed 1,000 times with replacement, in which some individuals can be sampled multiple times, while some may be sampled for 0 times. Then GWASs were conducted 1,000 times to see if there were still significant signals in the susceptibility region identified in our study. Our null hypothesis of bootstrap in our study is that more than 950 out of the 1,000 GWASs did not detect significant signals in the candidate region, which indicates that the fluctuation in the data structure of our experimental population has an effect on GWASs; in other words, the significant signals obtained in the GWASs were not accidental but were caused by differences in the genomes of the experimental individuals, which were reliable (Xu et al., 2019).

RESULTS

Phenotype and Genetic Parameter Statistics

Supplementary Figure S2 in **Supplementary File S1** show that three phenotypes follow the normal distribution. Descriptive statistics of the AFE, EP43w, and EP66w across the whole laying period are shown in **Table 1**. The mean value of AFE in this population was 136.95 days, which means that Shaoxing duck started laying eggs at about 20 weeks of age. Moreover, the mean values of EP43w and EP66w were 151.27 and 268.96, respectively. Estimates of SNP-based heritability as well as phenotypic correlations between AFE, EP43w, and EP66w are

TABLE 2 | Estimates of SNP-based heritability (on the diagonal) and of phenotypic correlations between traits (below the diagonal).

Traits	AFE	EP43w	EP66w
AFE	0.15		
EP43w	−0.73	0.20	
EP66w	−0.34	0.59	0.22

Note. AFE, age at first egg; EP43w, the egg number from onset of laying eggs to 43 weeks; EP66w, the egg number from onset of laying eggs to 66 weeks.

displayed in **Table 2**. The heritability was medium for all the three phenotypes, which was 0.15, 0.20, and 0.22 for AFE, EP43w, and EP66w, respectively. Genetic correlation analyses revealed that EP43w and EP66w were positively interrelated and were negatively interrelated with AFE.

Genome-Wide Association Study

After quality control, a total of 6,746,746 SNPs and 166 individuals were retained for further analyses. Association tests for AFE, EP43w, and EP66w were performed using a univariate linear model, and the threshold obtained by the naïve Bonferroni was 7.41E−09. The result showed that there was no SNP in the AFE and EP43w that surpassed this threshold, except for two SNPs on chromosome 29 that surpassed this threshold for EP66w (**Figures 1A–C**). It was easy to detect that the most significant sites appeared on chromosome 25 for AFE and EP43w, so we performed FDR correction on the *p*-values of those sites on chromosome 25, and all significantly associated loci that surpassed the FDR corrective threshold are shown in **Table 3**. In detail, we identified 12 SNPs that surpassed the FDR corrective genome-wide significance level for AFE (**Figure 1D**), and the most significantly associated SNP 25_4513397 (P_wald = 5.01E−08, Qvalue = 3.24E−03) was located at 4,513,397 bp within a 0.68-Mb region (4.44–5.12 Mb) on chromosome 25 (**Figure 2D**). We identified a total of 17 SNPs that surpassed the FDR corrective genome-wide significance level for EP43w (**Figure 1E**), and the most significantly associated SNP 25_3219815 (*p*-value = 2.91E−08, Qvalue = 1.881E−03) was located at 3,219,815 bp on chromosome 25 (**Figure 2E**). In addition, there was another QTL (4,442,034–4,513,397 bp) on chromosome 25 also associated with EP43w, and the most significantly associated SNP was 25_4442034 (P_wald = 4.05E−08, Qvalue = 1.309E−03) (**Figure 2E**). For the EP66w trait, we also identified 9 and 3 SNPs on chromosome 2 and chromosome 29 significantly associated with EP66w, respectively; the most associated SNP 2_129902811 (P_wald = 1.15E−08, Qvalue = 4.075E−03) and SNP 29_4481956 (P_wald = 2.39E−09, Qvalue = 2.535E−03) were located at 129,902,811 bp on chromosome 2 and 4,481,956 bp on chromosome 25, respectively (**Figure 2F**). In addition, to validate the possibility of spurious SNPs caused by population stratification, the Q-Q plots for these GWASs were explored (**Supplementary Figure S1**). The average inflation factors (λ) of the GWASs were 1.01, 1.02, and 1.01 in the three traits, indicating that population structures were properly corrected.

Post Genome-Wide Association Analysis

Through LD (Atzmon et al.) analysis, for AFE trait, we identified 23 SNPs that have strong LD status in the most significantly associated SNP 25_4513397 (**Figure 2A**, **Supplementary Table S4**), which were located within a 0.01-Mb region that spans from 4.511 to 4.521 Mb on chromosome 25. The candidate genes within the 0.01-Mb region included GRIK4 and ARHGEF12. For EP43w, we identified 25 SNPs that have strong LD status in the most significantly associated SNP 25_3219815 (**Figure 2B**, **Supplementary Table S4**), which were located within a 0.06-Mb region that spans from 3.186 to 3.247 Mb on chromosome 25. The candidate genes within the 0.06-Mb region involved five genes, including B3GAT1, VPS26B, ACAD8, THYN1, and NCAPD3. On chromosome 25, we also identified 8 SNPs that have strong LD status in the significantly associated SNP 25_4442034 (**Figure 2B**, **Supplementary Table S4**), which were located within a 0.02-Mb region that spans from 4.442 to 4.446 Mb. For the EP66w trait, we identified 318 SNPs that have strong LD status in the most significantly associated SNP 2_129902811 (**Figure 2C**, **Supplementary Table S4**), which were located within a 2.412-Mb region that spans from 127.497 to 129.910 Mb on chromosome 2. The candidate genes of EP66w within the 2.412 Mb involved six genes, including RALYL, LRRCC1, E2F5, RBIS, CA13, and CA2. We also identified 17 SNPs that have strong LD status in the most significantly associated SNP 29_4481956 (**Figure 2C**, **Supplementary Table S4**), which were located within a 0.355-Mb region that spans from 4.481 to 4.837 Mb on chromosome 29. The candidate genes within the 0.355-Mb region involved 13 genes, including DAZAP1, GAMT, NDUFS7, CIRBP, FAM174C, MIDN, STK11, SBNO2, POLR2E, ARHGAP45, GRIN3B, TMEM259, and WDR18. Overall, we identified a total of 26 candidate genes associated with the AFE, EP43w, and EP66w traits. Next, these genes were used to perform GO based on biological process analysis in DAVID (available at http://david.abcc.ncifcrf.gov/home.jsp), nine significant GO terms were identified (**Supplementary Figure S3**, **Supplementary Table S5**), and most genes are enriched in cytoplasm term and cytosol term.

Haplotype-Sharing Analysis

Through LD analysis, we obtained some corresponding CIs for AFE, EP43w, and EP66w, and then we performed a haplotype-sharing analysis of these intervals. The results are shown in **Figure 3A** and **Supplementary Table S6** in **Supplementary File S2**. We found that 184 sequences shared a type of haplotype for EP43w, defined as haplotype 1; the mean value of haplotype 1 with a corresponding phenotype was 148.86; the other haplotypes consisted of the remaining 149 sequences without any regularity, so we defined them as chaotic haplotypes, and the corresponding mean of those phenotypes was 154.22. Next, we carried out a *t*-test with haplotype 1 and chaotic haplotype (*p*-value = 0.0001), which indicated that haplotype 1 has a significant effect on EP43w. In addition, we also found four haplotypes in the CI (4,442,034 to 4,446,727 bp) that were related to EP43w (**Figure 3B** and **Supplementary Table S7** in **Supplementary File S2**) and named them haplotype 1,

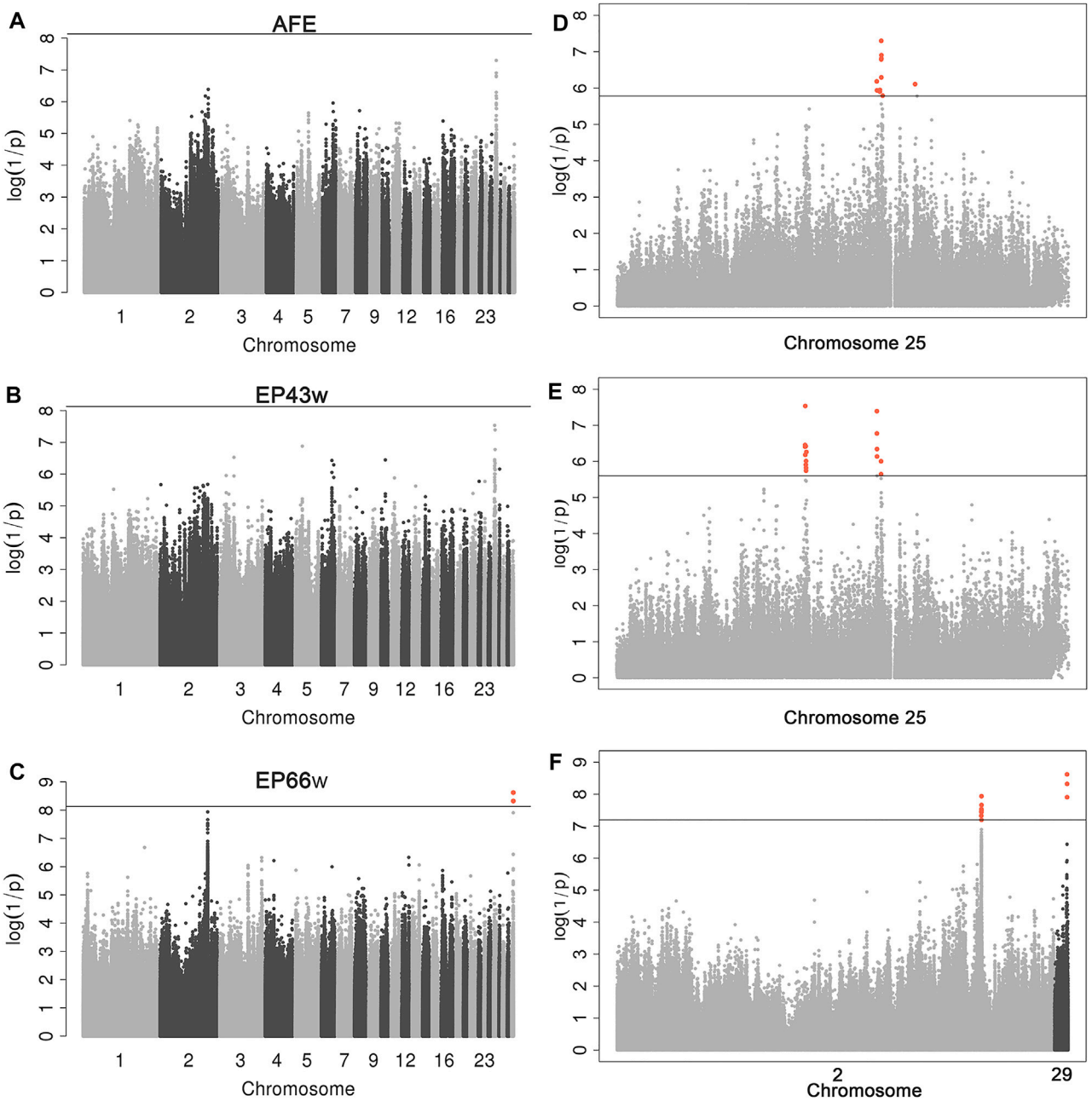


FIGURE 1 | Manhattan plots derived from GWASs for AFE, EP43w, and EP66w. Each dot on this figure corresponds to a SNP within the dataset, while the *y*-axis and *x*-axis represent the negative log10 *p*-value of the SNPs and the genomic positions separated by chromosomes, respectively. Black solid lines in panels **A–C** indicate the 5% genome-wide Bonferroni-corrected threshold; the tomato puree points represent SNPs that exceeded the chromosome-wide significance threshold. Black solid lines in panels **D–F** indicate the genome-wide FDR-corrected threshold; the tomato puree points represent SNPs that exceeded this threshold. GWASs, genome-wide association studies; AFE, age at first egg; EP43w, egg production at 43 weeks; EP66w, egg production at 66 weeks; SNP, single-nucleotide polymorphism; FDR, false discovery rate.

haplotype 2, haplotype 3, and haplotype 4, respectively, with the mean of 147.2, 149.9, 148.5, and 159.5 for the corresponding phenotypes. As there is no difference between haplotypes 1, 2, and 3, we merged those three haplotypes and did a *t*-test with haplotype 4, resulting in a *p*-value of 4.19E−05, which indicated that of the haplotypes, haplotype 4 has a significant effect of increasing EP43w. As the results show in **Figure 3C** and **Supplementary Table S8** in **Supplementary File S2**, we found that there were 220 sequences located on chromosome 29 that shared a type of haplotype for EP66w, defined as haplotype 1. The mean value of haplotype 1 with a corresponding phenotype was 272.89. The other haplotypes consisted of the remaining 112

TABLE 3 | Description of the significant SNPs associated with AFE, EP43w, and EP66w.

EP43w				AFE				EP66w			
Chr	Position	P_wald	Qvalue	Chr	Position	P_wald	Qvalue	Chr	Position	P_wald	Qvalue
25	3,219,815	2.91E–08	1.881E–03	25	4,513,397	5.01E–08	3.243E–03	29	4,481,956	2.39E–09	2.535E–03
25	4,442,034	4.05E–08	1.309E–03	25	4,516,366	1.24E–07	4.010E–03	29	4,500,604	4.75E–09	2.517E–03
25	4,442,632	1.68E–07	3.621E–03	25	4,515,630	1.53E–07	3.302E–03	2	129,902,811	1.15E–08	4.075E–03
25	3,216,505	3.52E–07	5.684E–03	25	4,513,382	1.62E–07	2.613E–03	29	4,500,595	1.24E–08	3.280E–03
25	3,227,771	3.80E–07	4.909E–03	25	4,514,932	5.06E–07	6.547E–03	2	129,903,599	2.17E–08	4.600E–03
25	3,216,680	3.91E–07	4.217E–03	25	4,436,853	6.52E–07	7.034E–03	2	129,877,347	2.94E–08	5.184E–03
25	4,443,950	4.56E–07	4.209E–03	25	5,092,232	7.77E–07	7.184E–03	2	129,903,026	3.22E–08	4.869E–03
25	3,238,808	5.49E–07	4.436E–03	25	4,490,115	1.12E–06	9.051E–03	2	129,836,874	3.51E–08	4.645E–03
25	3,220,324	6.57E–07	4.721E–03	25	4,441,061	1.15E–06	8.271E–03	2	129,877,317	3.61E–08	4.255E–03
25	4,444,559	7.30E–07	4.724E–03	25	4,489,078	1.25E–06	8.065E–03	2	129,877,399	3.61E–08	3.829E–03
25	3,233,091	9.80E–07	5.764E–03	25	4,537,557	1.61E–06	9.480E–03	2	129,826,588	4.70E–08	4.530E–03
25	4,513,397	9.92E–07	5.345E–03	25	5,126,926	1.65E–06	8.903E–03	2	129,903,609	6.34E–08	5.598E–03
25	3,229,361	1.25E–06	6.197E–03								
25	3,232,968	1.52E–06	7.017E–03								
25	3,232,878	1.81E–06	7.798E–03								
25	4,513,382	2.26E–06	9.150E–03								
25	4,444,987	2.51E–06	9.550E–03								

Note. Chr, chromosome number; position, base positions on the chromosome; P_wald, p*-value from the wald test; Qvalue,* p*-value corrected by FDR; SNPs, single-nucleotide polymorphisms; AFE, age at first egg; EP43w, the egg number from onset of laying eggs to 43 weeks; EP66w, the egg number from onset of laying eggs to 66 weeks; FDR, false discovery rate.*

sequences also without any regularity and are defined as chaotic haplotypes, with a corresponding mean of phenotypes of 260.4. The result of the *t*-test with haplotype 1 and chaotic haplotype (*p*-value = 2.8E−05) is indicative that haplotype 1 has a significant effect of increasing EP66w. For EP66w (**Figure 3D** and **Supplementary Table S9** in **Supplementary File S2**), another CI that spans from 127.497 to 129.910 Mb on chromosome 2 contained 318 loci, and we selected the loci with LD > 0.8 for haplotype-sharing analysis. The result showed that 299 sequences shared a type of haplotype (haplotype 1), the mean value of haplotype 1 with a corresponding phenotype was 271.94, and the other haplotypes consisting of the remaining 33 sequences were also defined as chaotic haplotypes, which correspond to the mean of those phenotypes at 241.87. Then we carried out a *t*-test with haplotype 1 and chaotic haplotype (*p*-value = 1.05E−09), which indicated that haplotype 1 has a significant effect on EP66w.

Bootstrap Test

Although these studies revealed some crucial discoveries, there were some limitations, such as the relatively small number of samples in our experimental population. Therefore, we herein carried out a bootstrap test to verify the reliability of GWASs in our study. For trait EP43w, there are 985 of the 1,000 GWASs that did not detect significant signals (P_wald < 1.81E−06) in the interval from 3,216,505 to 3,238,808 bp on chromosome 25, and there are 931 of the 1,000 GWASs that did not detect significant signals in the interval from 4,442,034 to 4,513,397 bp on chromosome 25. For trait EP66w, there are 987 of the 1,000 GWASs that did not detect significant signals (P_wald < 6.34E−08) in the interval from 129,826,588 to 129,903,609 bp on chromosome 2, and there are 992 of the 1,000 GWASs that did not detect significant signals (P_wald < 1.24E−08) in the interval from 4,481,956 to 4500595 bp on chromosome 29. These results indicated that the fluctuation in the data structure of our experimental population has an effect on GWASs (FDR < 0.05); in other words, the significant signals obtained in our GWAS were not accidental but were caused by differences in the genomes of the experimental individuals, which were reliable.

DISCUSSION

Egg production is an important economic trait. So far, many studies have focused on the genetic determinants of AFE, EP43w, and EP66w in chicken and have reported some candidate QTLs and genes (Liu et al., 2011; Goraga et al., 2012; Wolc et al., 2014; Yuan et al., 2015; Kudinov et al., 2019; Liu et al., 2019). However, the molecular genetic research that affects egg-laying performance in laying ducks is still limited, with only a few candidate gene studies. GWAS has become a powerful approach for genetic dissection of trait loci along with the completion of genome sequencing and the development of a high-density SNP array. In our study, we performed a GWAS for AFE, EP43w, and EP66w using a univariate linear mixed model. This is the first GWAS that used the whole-genome sequencing in a Shaoxing pure line population across the whole laying period.

Genetic parameter estimates show that AFE, EP43w, and EP66w are medium heritable traits, which approximately coincided with the report by Chen et al. (Chen and Tan 1996). Our research is the first report of heritability estimates of egg production in laying ducks using the whole-genome sequencing, which can provide some reference for subsequent studies on egg production in laying ducks.

We conducted a GWAS in Shaoxing duck population and provided strong evidence of the association of SNPs with 3 traits of egg production. There is an LD between the marker SNP and the causative variation within or near genes, as most SNPs found at genome-wide significance level in our study are within the

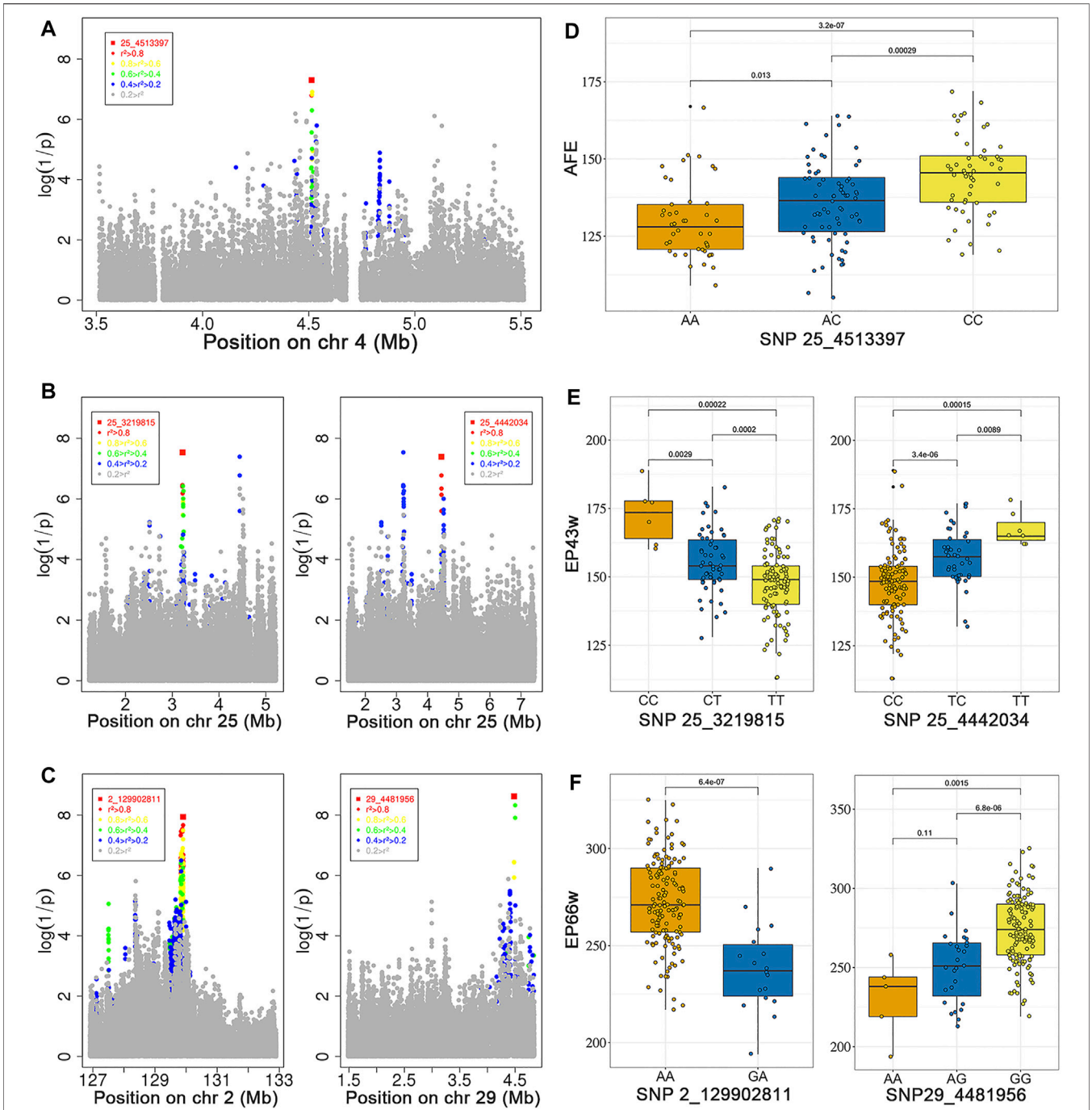


FIGURE 2 | Regional plots for the strongly associated region in the GWAS for AFE **(A)**, EP43w **(B),** and EP66w **(C)**. The horizontal coordinates indicate the strongly associated region, and the vertical coordinates indicate the *p*-values; the color of each locus indicates the LD status in the most significantly associated locus. Distribution of phenotypic values for the three SNP genotypes most associated with AFE **(D)**, EP43w **(E),** and EP66w **(F)**. GWAS, genome-wide association study; AFE, age at first egg; EP43w, egg production at 43 weeks; EP66w, egg production at 66 weeks; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

known genes. Identifications of these loci may provide new insights into the genetic basics of egg production traits, though the characteristics and functions of these genes have not been studied in depth.

Number of eggs and AFE are two important production traits in laying ducks, and producing laying duck with earlier sexual maturity and a higher rate has always been the goal of laying duck breeding. Our study indicated that these reproductive traits are sex-limited and have low-to-moderate heritability, indicating that they can be genetically improved by marker-assisted selection and genomic selection. In this study, we found two candidate genes that affect AFE, including GRIK4 and

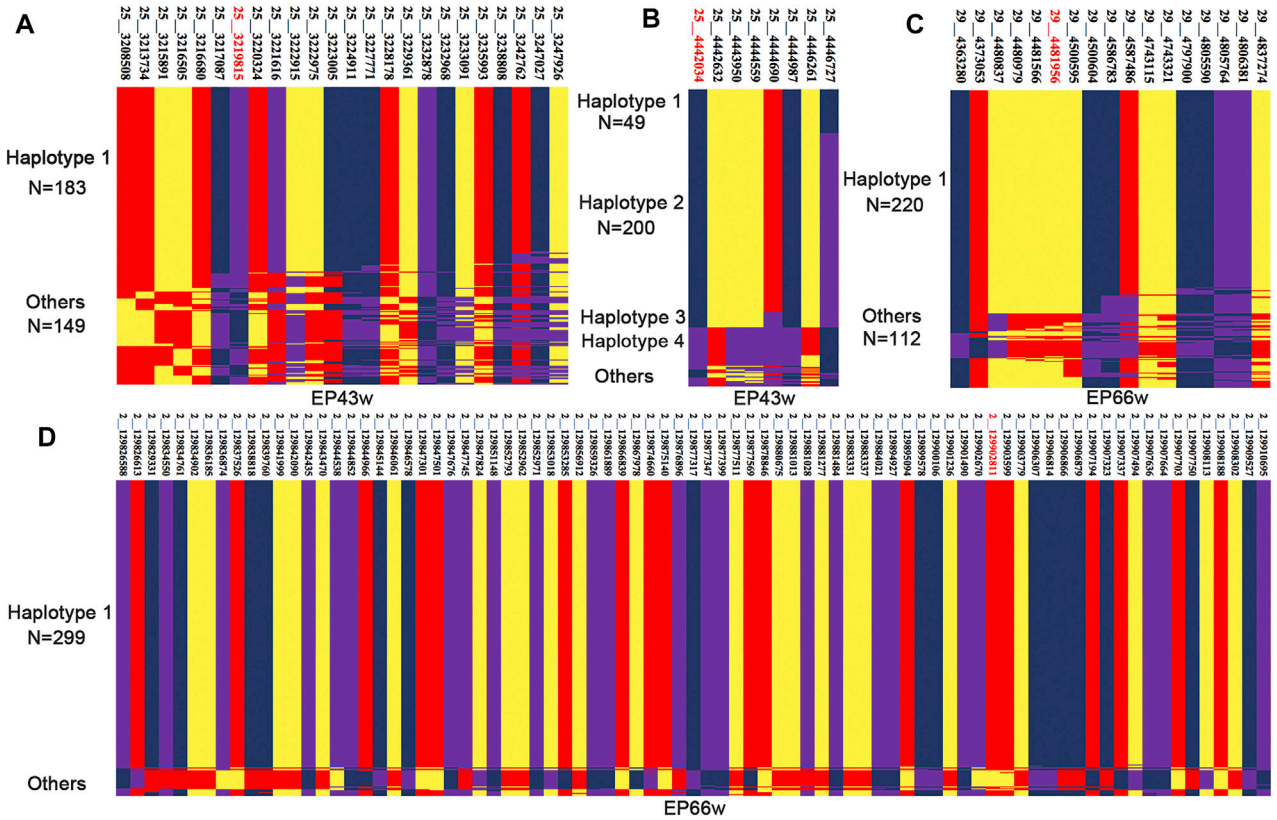


FIGURE 3 | The haplotype sharing within the CI for EP43w and EP66w. EP43w, egg production at 43 weeks; EP66w, egg production at 66 weeks.

ARHGEF12, and we found five candidate genes that affect EP43w, including B3GAT1, VPS26B, ACAD8, THYN1, and NCAPD3. ACADs are a family of mitochondrial flavoenzymes that catalyze the dehydrogenation steps of the α- and β-oxidation processes, which are related to fatty acid β-oxidation. Lv et al. have found that dietary genistein supplementation in feed inhibited fatty acid synthesis and enhanced β-oxidation in the livers of layers with fatty liver syndrome through the PPAR–ACAD pathways, thereby alleviating fat deposition and lipid metabolism disorder, resulting in significant improvement in the laying rate poultry (Lv et al., 2018). Yuan et al. found THYN1 was associated with immune and cytokines, which played essential modulatory roles in the regulation of ovarian function (Onagbesan et al., 2009; Yuan et al., 2015). We found six genes located in CHROMOSOME 2 that affect EP66w, including RALYL, LRRCC1, E2F5, RBIS, CA13, and CA2. Carbonic anhydrase II (CA2) is a widespread zinc metalloenzyme from the carbonic anhydrase family and is essential for osteoclast activity, hydration of carbon dioxide, and pH balance (Roth et al., 1992; Geers and Gros, 2000). Nys and de Laage reported that the level of carbonic anhydrase is lower in the uterus and duodenum of hens laying soft-shelled eggs (Nys and de Laage, 1984). Some studies have proposed that disrupted carbonic anhydrase expression and distribution are involved in the mechanism of estrogen-induced eggshell thinning (Holm et al., 2001; Berg et al., 2004). Dunn et al. reported that CA2 gene polymorphism is associated with chicken egg shape (Dunn et al., 2009). Especially, Chang et al. found that CA2 is one of the differentially expressed transcripts in the duck isthmus epithelium during the egg formation period, and they confirmed that some SNPs in the 3′-UTR of the CA2 gene in Tsaiya ducks are associated with egg reproduction traits (Chang et al., 2013). We found 13 genes located in CHROMOSOME 29 that affect EP66w, including DAZAP1, GAMT, NDUFS7, CIRBP, FAM174C, MIDN, STK11, SBNO2, POLR2E, ARHGAP45, GRIN3B, TMEM259, and WDR18. Guanidinoacetate *N*-methyltransferase (GAMT) has been shown to be associated with the reproductive system and development, which implies that GAMT may be a candidate gene underlying egg production traits (Singh et al., 2019). In addition, our study also found some haplotypes that were significantly associated with these three traits, which can be helpful to improve egg production performance in laying duck based on breeding.

The relatively small number of samples in our experimental population is a limitation of this study. Therefore, we refer to a method called the bootstrap test to verify the reliability of GWASs in this study. The result showed that significant signals obtained in our GWASs were not accidental and were reliable. We have uploaded this method to the GitHub website, and users can access this method at https://github.com/xuwenwu24/Bootstrap-test.

CONCLUSION

In summary, this study demonstrates that AFE, EP43w, and EP66w have medium heritability, and there were strong correlations between them. We have located some significant confidence regions for those traits, and some genes, such as GRIK4 ARHGEF12, ACAD8, THYN1, CA2, and GAMT, may be the putative candidate genes underlying this interval based on its biochemical and physiological functions. In addition, our study also found some haplotypes that were significantly associated with these three traits. Post-study can identify causal mutations by enriching markers within the identified intervals and functional studies on related genes.

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, PRJCA005720.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of 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

LZ and ZT conceived and designed the experiments. XW and WZ analyzed the data. CQ, QQ, LY, TY, CL, TH, LF, LQ, SD, and TR contributed to materials and analysis tools. XW wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Waterfowl Industry Technology System (No. CARS-43-02) and Zhejiang Major Scientific and Technological project of agricultural (livestock) breeding (No. 2016C02054-12). The founder LLZ had a role in the design of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.828884/full#supplementary-material

REFERENCES

Atzmon, G., Blum, S., Feldman, M., Cahaner, A., Lavi, U., and Hillel, J. (2008). QTLs Detected in a Multigenerational Resource Chicken Population. *J. Hered.* 99 (5), 528–538. doi:10.1093/jhered/esn030

Berg, C., Blomqvist, A., Holm, L., Brandt, I., Brunström, B., and Ridderstråle, Y. (2004). Embryonic Exposure to Oestrogen Causes Eggshell Thinning and Altered Shell Gland Carbonic Anhydrase Expression in the Domestic Hen. *Reproduction* 128 (4), 455–461. doi:10.1530/rep.1.00211

Chang, M.-T., Cheng, Y.-S., and Huang, M.-C. (2013). Novel Genetic Markers of the Carbonic Anhydrase II Gene Associated with Egg Production and Reproduction Traits in Tsaiya Ducks. *Reprod. Domest. Anim.* 48 (1), 98–104. doi:10.1111/j.1439-0531.2012.02038.x

Chen, H., and Tan, J. (1996). Estimation on Major Economic Character Hereditary Parameter of Putian Black Duck (Egg Duck). *J. Fujian Acad. Agric. Sci.* 9 (4), 35–38. doi:10.2174/1872208313666190404101336

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6 (2), 80–92. doi:10.4161/fly.19695

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data. *Nat. Genet.* 43 (5), 491–498. doi:10.1038/ng.806

Dunn, I. C., Joseph, N. T., Bain, M., Edmond, A., Wilson, P. W., Milona, P., et al. (2009). Polymorphisms in Eggshell Organic Matrix Genes Are Associated with Eggshell Quality Measurements in Pedigree Rhode Island Red Hens. *Anim. Genet.* 40 (1), 110–114. doi:10.1111/j.1365-2052.2008.01794.x

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap [M]*. Boca Raton, Florida: CRC press.

Geers, C., and Gros, G. (2000). Carbon Dioxide Transport and Carbonic Anhydrase in Blood and Muscle. *Physiol. Rev.* 80 (2), 681–715. doi:10.1152/physrev.2000.80.2.681

Goraga, Z. S., Nassar, M. K., and Brockmann, G. A. (2012). Quantitative Trait Loci Segregating in Crosses between New Hampshire and White Leghorn Chicken Lines: I. Egg Production Traits. *Anim. Genet.* 43 (2), 183–189. doi:10.1111/j.1365-2052.2011.02233.x

Goto, T., Ishikawa, A., Onitsuka, S., Goto, N., Fujikawa, Y., Umino, T., et al. (2011). Mapping Quantitative Trait Loci for Egg Production Traits in an F2 Intercross of Oh-Shamo and White Leghorn Chickens. *Anim. Genet.* 42 (6), 634–641. doi:10.1111/j.1365-2052.2011.02190.x

Holm, L., Berg, C., Brunström, B., Ridderstråle, Y., and Brandt, I. (2001). Disrupted Carbonic Anhydrase Distribution in the Avian Shell Gland Following in Ovo Exposure to Estrogen. *Arch. Toxicol.* 75 (6), 362–368. doi:10.1007/s002040100241

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., et al. (2013). The Duck Genome and Transcriptome Provide Insight into an Avian Influenza Virus Reservoir Species. *Nat. Genet.* 45 (7), 776–783. doi:10.1038/ng.2657

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., et al. (2005). Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* 308 (5720), 385–389. doi:10.1126/science.1109557

Kudinov, A. A., Dementieva, N. V., Mitrofanova, O. V., Stanishevskaya, O. I., Fedorova, E. S., Larkina, T. A., et al. (2019). Genome-wide Association Studies Targeting the Yield of Extraembryonic Fluid and Production Traits in Russian White Chickens. *BMC Genomics* 20 (1), 270. doi:10.1186/s12864-019-5605-5

Kwon, J. M., and Goate, A. M. (2000). The Candidate Gene Approach. *Alcohol. Res. Health* 24 (3), 164–168. Available at: https://permanent.fdlp.gov/lps1441/lps1441/pubs.niaaa.nih.gov/publications/arh24-3/164-168.pdf.

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

Liu, W., Li, D., Liu, J., Chen, S., Qu, L., Zheng, J., et al. (2011). A Genome-wide SNP Scan Reveals Novel Loci for Egg Production and Quality Traits in white Leghorn and Brown-Egg dwarf Layers. *PLoS One* 6 (12), e28600. doi:10.1371/journal.pone.0028600

Liu, Z., Yang, N., Yan, Y., Li, G., Liu, A., Wu, G., et al. (2019). Genome-wide Association Analysis of Egg Production Performance in Chickens across the Whole Laying Period. *BMC Genet.* 20 (1), 67. doi:10.1186/s12863-019-0771-7

Lv, Z., Xing, K., Li, G., Liu, D., and Guo, Y. (2018). Dietary Genistein Alleviates Lipid Metabolism Disorder and Inflammatory Response in Laying Hens with Fatty Liver Syndrome. *Front. Physiol.* 9, 1493. doi:10.3389/fphys.2018.01493

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157 (4), 1819–1829. doi:10.1093/genetics/157.4.1819

Muir, W. M. (2007). Comparison of Genomic and Traditional BLUP-Estimated Breeding Value Accuracy and Selection Response under Alternative Trait and Genomic Parameters. *J. Anim. Breed. Genet.* 124 (6), 342–355. doi:10.1111/j.1439-0388.2007.00700.x

Nys, Y., and de Laage, X. (1984). Effects of Suppression of Eggshell Calcification and of 1,25(OH)2D3 on Mg2+ Ca2+ and Mg2+HCO3- ATPase, Alkaline Phosphatase, Carbonic Anhydrase and CaBP Levels-I. The Laying Hen Uterus. *Comp. Biochem. Physiol. A: Physiol.* 78 (4), 833–838. doi:10.1016/0300-9629(84)90642-x

Onagbesan, O., Bruggeman, V., and Decuypere, E. (2009). Intra-ovarian Growth Factors Regulating Ovarian Function in Avian Species: a Review. *Anim. Reprod. Sci.* 111 (2-4), 121–140. doi:10.1016/j.anireprosci.2008.09.017

Pearson, T. A., and Manolio, T. A. (2008). How to Interpret a Genome-wide Association Study. *JAMA* 299 (11), 1335–1344. doi:10.1001/jama.299.11.1335

Roth, D. E., Venta, P. J., Tashian, R. E., and Sly, W. S. (1992). Molecular Basis of Human Carbonic Anhydrase II Deficiency. *Proc. Natl. Acad. Sci.* 89 (5), 1804–1808. doi:10.1073/pnas.89.5.1804

Sasaki, O., Odawara, S., Takahashi, H., Nirasawa, K., Oyamada, Y., Yamamoto, R., et al. (2004). Genetic Mapping of Quantitative Trait Loci Affecting Body Weight, Egg Character and Egg Production in F2 Intercross Chickens. *Anim. Genet.* 35 (3), 188–194. doi:10.1111/j.1365-2052.2004.01133.x

Scheet, P., and Stephens, M. (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *Am. J. Hum. Genet.* 78 (4), 629–644. doi:10.1086/502802

Schreiweis, M. A., Hester, P. Y., Settar, P., and Moody, D. E. (2006). Identification of Quantitative Trait Loci Associated with Egg Quality, Egg Production, and Body Weight in an F2 Resource Population of Chickens1. *Anim. Genet.* 37 (2), 106–112. doi:10.1111/j.1365-2052.2005.01394.x

Singh, V., Bala, R., Chakraborty, A., Rajender, S., Trivedi, S., and Singh, K. (2019). Duplications in 19p13.3 Are Associated with Male Infertility. *J. Assist. Reprod. Genet.* 36 (10), 2171–2179. doi:10.1007/s10815-019-01547-1

Tuiskula-Haavisto, M., Honkatukia, M., Vilkki, J., de Koning, D., Schulman, N., and Maki-Tanila, A. (2002). Mapping of Quantitative Trait Loci Affecting Quality and Production Traits in Egg Layers. *Poult. Sci.* 81 (7), 919–927. doi:10.1093/ps/81.7.919

Wang, C., Li, S. J., Li, C., Yu, G. H., Feng, Y. P., Peng, X. L., et al. (2013). Molecular Cloning, Expression and Association Study with Reproductive Traits of the duckLRP8gene. *Br. Poult. Sci.* 54 (5), 567–574. doi:10.1080/00071668.2013.819488

Wolc, A., Arango, J., Jankowski, T., Dunn, I., Settar, P., Fulton, J. E., et al. (2014). Genome-wide Association Study for Egg Production and Quality in Layer Chickens. *J. Anim. Breed. Genet.* 131 (3), 173–182. doi:10.1111/jbg.12086

Xu, H., Zeng, H., Luo, C., Zhang, D., Wang, Q., Sun, L., et al. (2011). Genetic Effects of Polymorphisms in Candidate Genes and the QTL Region on Chicken Age at First Egg. *BMC Genet.* 12, 33. doi:10.1186/1471-2156-12-33

Xu, W., Chen, D., Yan, G., Xiao, S., Huang, T., Zhang, Z., et al. (2019). Rediscover and Refine QTLs for Pig Scrotal Hernia by Increasing a Specially Designed F3 Population and Using Whole-Genome Sequence Imputation Technology. *Front. Genet.* 10, 890. doi:10.3389/fgene.2019.00890

Yoav, B., and Daniel, Y. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* 29 (4), 1165–1188. doi:10.1002/bimj.200510311

Yuan, J., Sun, C., Dou, T., Yi, G., Qu, L., Qu, L., et al. (2015). Identification of Promising Mutants Associated with Egg Production Traits Revealed by Genome-wide Association Study. *PLoS One* 10 (10), e0140615. doi:10.1371/journal.pone.0140615

Zhou, X., and Stephens, M. (2012). Genome-wide Efficient Mixed-Model Analysis for Association Studies. *Nat. Genet.* 44 (7), 821–824. doi:10.1038/ng.2310

Conflict of Interest: Authors YQ, QL, JT, and CL are all employed by Hubei Shendan Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships.

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

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fvets.2022.816241

Comparative Transcriptomics Reveals the Key lncRNA and mRNA of Sunite Sheep Adrenal Gland Affecting Seasonal Reproduction

Xiaolong Du[1†], *Xiaoyun He*[1†], *Qiuyue Liu*[1], *Ran Di*[1], *Qingqing Liu*[1,2] *and Mingxing Chu*[1*]

[1] *Key Laboratory of Animal Genetics, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China,* [2] *College of Animal Science and Technology, Anhui Agricultural University, Hefei, China*

OPEN ACCESS

Edited by:
Hasan Riaz,
COMSATS Institute of Information Technology, Pakistan

Reviewed by:
Xin Qi,
Ocean University of China, China
Awais Ihsan,
COMSATS University Islamabad, Sahiwal Campus, Pakistan

***Correspondence:**
Mingxing Chu
mxchu@263.net

[†]*These authors have contributed equally to this work*

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

Received: *16 November 2021*
Accepted: *03 March 2022*
Published: *08 April 2022*

Citation:
Du X, He X, Liu Q, Di R, Liu Q and Chu M (2022) Comparative Transcriptomics Reveals the Key lncRNA and mRNA of Sunite Sheep Adrenal Gland Affecting Seasonal Reproduction. Front. Vet. Sci. 9:816241. doi: 10.3389/fvets.2022.816241

The hypothalamic–pituitary–adrenal (HPA) axis plays an important role in the growth and development of mammals. Recently, lncRNA transcripts have emerged as an area of importance in sheep photoperiod and seasonal estrus studies. This research aims to identify lncRNA and mRNA that are differentially expressed in the sheep adrenal gland in long (LP) or short (SP) photoperiods using transcriptome sequencing and bioinformatics analysis based on the OVX + E_2 (Bilateral ovariectomy and estradiol-implanted) model. We found significant differences in the expression of lncRNAs in LP42 (where LP is for 42 days) vs. SP-LP42 (where SP is for 42 days followed by LP for 42 days) ($n = 304$), SP42 (where SP is for 42 days) vs. SP-LP42 ($n = 1{,}110$) and SP42 vs. LP42 ($n = 928$). Cluster analysis and enrichment analysis identified SP42 vs. LP42 as a comparable group of interest and found the following candidate genes related to reproductive phenotype: *FGF16*, *PLGF*, *CDKN1A*, *SEMA7A*, *EDG1*, *CACNA1C* and *ADCY5*. *FGF16* (Up-regulated lncRNA MSTRG.242136 and MSTRG.236582) is the only up-regulated gene that is closely related to oocyte maturation. However, *EDG1* (Down-regulated lncRNA MSTRG.43609) and CACNA1C may be related to precocious puberty in sheep. *PLGF* (Down-regulated lncRNA MSTRG.146618 and MSTRG.247208) and *CDKN1A* (Up-regulated lncRNA MSTRG.203610 and MSTRG.129663) are involved in the growth and differentiation of placental and retinal vessels, and *SEMA7A* (Up-regulated lncRNA MSTRG.250579) is essential for the development of gonadotropin-releasing hormone (GnRH) neurons. These results identify novel candidate genes that may regulate sheep seasonality and may lead to new methods for the management of sheep reproduction. This study provides a basis for further explanation of the basic molecular mechanism of the adrenal gland, but also provides a new idea for a comprehensive understanding of seasonal estrus characteristics in Sunite sheep.

Keywords: HPA axis, seasonality, photoperiod, candidate gene, sheep

INTRODUCTION

Animals that show seasonal reproduction patterns only mate at certain times as their reproductive cycles start and stop based on the season (1, 2). Sheep are seasonal breeders and are often used as a model species to study the effect of photoperiod on reproductive function (2–4). The reproductive endocrine axis of ewes is affected by variations in photoperiod. Ewes transition from an estrus state to an anestrus state from spring to autumn (5, 6). Light affects the secretion of melatonin, which leads to changes in the circadian rhythm of seasonal reproduction of animals (7). Melatonin is produced by the pineal gland, which then acts on the hypothalamus, affecting sheep reproduction through the hypothalamic–pituitary–gonadal (HPG) axis (8–11). The HPG axis and HPA axis are closely related and influence each other. For example, the PVN (paraventricular nucleus) is stimulated to secrete corticotropin-releasing hormone (CRH) which then activates the release of adrenocorticotropic hormone (ACTH) from the pituitary. The ACTH, in turn, stimulates secrete of cortisol from the adrenals which then provide negative feedback back to the brain in a classic homeostatic feedback loop to fine-tune HPA axis signaling (12). Vast quantities of studies show that basal cortisol levels are higher in females than males and the capacity of glucocorticoid secrete was higher in females, suggesting that E_2 (17β-estradiol) maybe increases HPA axis reactivity (13).

Estrogen is one of the most important hormones in sheep reproduction, especially in anestrus animals. Estrogen negatively regulates the neuroendocrine circuit, affecting the secretion of GnRH (14, 15). Meanwhile, the E_2-induced surge pattern of luteinizing hormone (LH) and GnRH secretion that conducts ovulation in females, is assailable to the effects of cortisol (12). However, the exact molecular mechanism is not clear. A previous study by Luo et al. (16) found exogenous cortisol treatment of gonad-intact female mice restrained cyclicity in diestrus. Ovariectomy (OVX) female mice were treated with an LH surge-inducing E_2 implant, as well as a cortisol or cholesterol (control) pellet, and detected two days later for LH levels on the prospective LH surge. All cholesterol-treated females showed a clear LH surge, whereas LH levels were undetectable in cortisol-treated females (16). Many experiments have shown that glucocorticoids can affect the related function of LH. Such as cortisol after infusion of encephalocoele suppresses LH pulse amplitude in ovariectomized ewes (17). At present, most researchers use hypothalamic–pituitary disconnection (HPD) model to study the effect of photoperiod on sheep reproduction, and it has been proved that prolactin is a key hormone involved in the seasonal reproduction of sheep (18–20). It has also been proved that OVX + E_2 model is also a classical model for the study of photoperiod regulation and hypothalamic function (21, 22).

The rapid development of RNA-seq technology has improved the efficiency of animal molecular genetics and breeding. Long-stranded, non-coding RNA (lncRNA) is a non-coding RNA with a length of more than 200 bp (23, 24). Studies have shown that lncRNA regulates many biological functions, including, dose compensation effect, epigenetics and cell differentiation (24). The topic has become a research hotspot across multiple scientific disciplines, and many lncRNAs have been associated with animal reproduction. For example, several lncRNAs have been associated with STH (Small-tailed Han sheep) fertility (25) and adolescent development in the hypothalamus of goats (26, 27). Moreover, analysis of the hypophysis of Hu sheep with high and low fertility identified 57 differentially expressed lncRNAs (28). These studies show that lncRNAs in the pituitary, and ovaries, of sheep have regulatory functions in reproduction (29). The adrenal gland influences reproduction in sheep (30–33), however, few studies have assessed the function of lncRNAs in this organ.

In this study we analyze the key candidate lncRNAs and mRNA in the HPA axis that affects seasonal reproduction of Sunite ewes through transcriptome sequencing of the adrenal gland. This provides a new perspective for the study of sheep seasonal reproduction.

MATERIALS AND METHODS

Ethics Statement and OVX + E_2 Model Building

Ethics approval (No. IAS2018–3) was granted by the Animal Ethics Committee of the Institute of Animal Sciences of Chinese Academy of Agricultural Sciences (IAS-CAAS) (Beijing, China). Nine non-pregnant adult Sunite ewes (aged 2–3 years old; weight 30–40 kg), which were randomly selected from a farm in Bayan Nur City (40°75′north latitude), Inner Mongolia Autonomous Region, China, were used for the study. The ovaries of these animals were removed by surgery, and an estrogen silicone tube was implanted subcutaneously in the neck of the sheep, as described previously (34–36). The ewes were randomly divided into three groups: SP42 (short photoperiod for 42 days; $n = 3$), LP42 (long photoperiod for 42 days; $n = 3$) and SP-LP42 (short photoperiod for 42 days followed by a long photoperiod for 42 days; $n = 3$). The conditions for the long photoperiod were 16 h of light per day and 8 h without light. The conditions for the short photoperiod were 8 h of light exposure and 16 h without light exposure. All sheep had free access to water and feed in an enclosed climate control chamber with only artificial light sources.

Tissues Acquisition and Sequencing

Adrenal gland tissue from euthanized ewes was quickly preserved in liquid nitrogen with tweezers. The stored tissues were used for RNA extraction with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instruction. The purity of the RNA samples was detected by a Nano Photometer® spectrophotometer (IMPLEN, Westlake Village, CA, USA). A

Abbreviations: lncRNA, Long non-coding RNA; OVX + E_2, ovariectomy and estrogen was embedded in the epidermis; HPA, the hypothalamic–pituitary–adrenal; mRNA, message RNA; LP, long photoperiod; SP, short photoperiod; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LP42, the LP lasts for 42 days; SP-LP42, the SP lasts for 42 days, and then convert it to LP for 42 days; SP42, the SP lasts for 42 days; *FGF16*, fibroblast growth factor 16; *PLGF*, placental growth factor; *EDG1*, Endothelial differentiation gene1; MVD, microvessel density; *S1PR1*, sphingosine 1–phosphate receptor 1; *SEMA7A*, semaphorin 7; *CACNA1C*, voltage-dependent calcium channel L type alpha-1C; *ADCY5*, adenylate cyclase 5.

Qubit® 3.0 RNA Assay kits (Life Technologies, CA, USA) and RNA Nano 6000 Assay (Agilent Technologies, CA, USA) were used to determine the integrity and concentration of RNA samples. The RNA integrity number (RIN) value of all samples being greater than seven.

The lncRNA library was constructed with 3 μg of high-quality RNA using the NEB Next Ultra Directional RNA Library Prep Kit (NEB, Ispawich, USA) for Illumina, according to the manufacturer's instructions. During this process, Ribo-Zero™ GoldKits (TEANGEN, Beijing, China) were used to remove rRNA. In addition, we used the UNG enzyme to degrade the second strand of U-containing cDNA and performed PCR amplification to obtain the RNA library, RNA-sequencing libraries were generated by paired-end (PE150) sequencing. The RNA library was then sequenced at a concentration of 1 ng/μL RNA using Hiseq 2500 (Illumina, San Diego, CA, United States). All sequencing data was outsourced to Annoroad Gene Technology Co., Ltd. (Beijing, China).

Data Quality Control and Transcriptome Assembly

Bcl2fastq (v2.17.1.14) is used to process the offline data and convert the original image file into raw sequencing reads on base calling, that was raw read. Clean reads were acquired using in-house Perl script made by Annoroad Genentech Co., Ltd. (Beijing, China) from the raw reads through the removal of: reads with adaptor contamination (i.e., adaptor reads with more than five contaminated bases), low-quality reads (i.e., more than 50% of the bases in the reading have a mass Phred Quality Score of $q \leq 19$), reads with a rate of N > 5% (i.e., for double-end sequencing, if one-end sequencing does not meet the above requirements, the reads of both ends are removed), and those that matched with ribosomal RNA. We used the *Ovis aries* reference genome (Oar_v4.0), and the genome annotation file from ENSEMBL. Clean reads were then mapped to the reference genome using HiSAT2 (v2.0.5) (37) and StringTie (v1.3.2d) was used to assemble the transcripts (38). HiSAT2 was run with "-rna—strandness RF" and "-dta -t -p 4," String Tie with "-G ref.gtf -rf—1," and the other parameters were set as the default.

lncRNAs and mRNAs Identification and Differential Expression Analysis

Novel lncRNAs transcripts were identified on the following conditions: its length is ≥200 bp, the number of exons is ≥2, and its reads coverage is >5. And remove the known mRNA and other non-coding RNA of the species. Importantly, the coding-non-coding index (CNCI) (39), the coding potential calculator (CPC) (40), the protein families database (PFAM) (41), and the coding potential assessment tool (CPAT) (42) software was used to determine if the transcripts had coding potential and whether they were new transcripts. CNCI was run with "–score 0 –length 199—exon_num 2" with the other parameters set as the default. In both CNCI and CPC, a score <0 was considered to indicate that the lncRNA could be defined as a non-coding RNA. Pfam was run with "minimum protein length: 60" and the other parameters set as the default. CPAT (v1.2.1) was used to screen the coding RNAs by constructing a logistic regression model and calculating Fickett score and Hexamer score, which were based on open reading frame (ORF) length and coverage, respectively.

We used the HTSeq Python package (v0.6.1) to calculate read counts, HTSeq was run with "-I gene_id -f bam -s" and "reverse -a 10 -q" with the other parameters set as the default. DESeq (43) was then applied to identify the differential expression of the lncRNAs based on the normalized counts by using three comparisons: SP42 vs. LP42?SP42 vs. SP-LP42 and LP42 vs. SP-LP42. In addition, $|\text{Log2Ratio}| \geq 1$ and $q < 0.05$ was considered to be screening threshold of significantly differential expression. The fragments per kilobase per million mapped reads (FPKM) were calculated to represent the expression levels of the lncRNAs and mRNAs (44). Based on the log2 (FPKM) value of mRNA and lncRNA, clustering analysis was performed using pheatmap (v1.0.2) to explore the similarities and analyze the relationships between the different libraries (45). The analysis consisted of Pearson's correlation and Euclidean distance methods.

Target Gene Prediction of lncRNAs and Gene Enrichment Analysis

To better understand the function of differentially expressed lncRNAs in SP42 vs. LP42, SP42 vs. SP-LP42 and LP42 vs. SP-LP42 we carried out target gene predictions. The target genes can be divided into cis-targets and trans-targets based on the distances and expressions correlation of lncRNAs and protein-coding genes. When the expression quantity correlation coefficient of a lncRNA, and its corresponding target mRNA, was ≥ 0.95 it was considered to be a potential trans-target. If the lncRNAs were located < 50 kb from nearby genes we assigned cis-targets function to them (24).

We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses by using the clusterProfiler package (v3.16.0) to clarify the potential roles of the targeted genes of differentially expressed lncRNAs. The hypergeometric test method was applied to assess significantly enriched GO terms and KEGG pathways. Those with false discovery rate (FDR) < 0.1 and $q < 0.05$, were considered to be significantly enriched.

Construction of Integral lncRNA–mRNA Interaction Networks

The regulatory network analysis of differentially expressed lncRNAs, and target genes, was drawn according to the relationship between the differentially expressed lncRNAs and mRNA genes, and the genes predicted by cis- and trans-targets of lncRNAs using Cytoscape software.

Data Validation

Transcripts ($n = 8$) were randomly selected and the primers were designed by primer 5.0 software. The designed primers were synthesized by Beijing Tianyi Huiyuan Biological Technology Co., Ltd. The qPCR reaction conditions were as follows: 95°C for 15 min, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. The data obtained from the qPCR reaction was evaluated using the $2^{-\Delta\Delta Ct}$ method and statistically analyzed using a one-way analysis of variance in the SPSS20.0. The results are presented as means ± standard deviation. $p < 0.05$ was considered statistically significant.

TABLE 1 | Summary of the mapping data from the adrenal gland tissues.

Items	Raw reads number	Clean reads number	Mapping rate	Clean Q30 bases rate (%)	Multimap rate
SP42A1	129,756,938	125,490,436	95.46%	94.07%	8.73%
SP42A2	113,449,408	110,469,132	94.98%	94.07%	6.43%
SP42A3	118,849,666	115,155,856	95.25%	94.03%	8.07%
LP42A1	115,231,936	111,648,952	94.94%	94.22%	7.50%
LP42A2	106,757,808	101,268,666	94.35%	94.53%	6.13%
LP42A3	115,480,206	111,612,316	95.27%	94.03%	8.73%
SPLP42A1	124,707,824	121,476,492	94.71%	94.19%	6.65%
SPLP42A2	123,408,246	120,023,448	95.13%	94.14%	7.05%
SPLP42A3	117,474,594	111,722,560	94.96%	93.70%	6.66%

LP42, and SP42 represent a long light period every day for 42 days and a short light period every day for 42 days. SP-LP42 means short light period every day for 42 days, followed by a long light period for 42 days. A represents the adrenal gland in the sample ID.

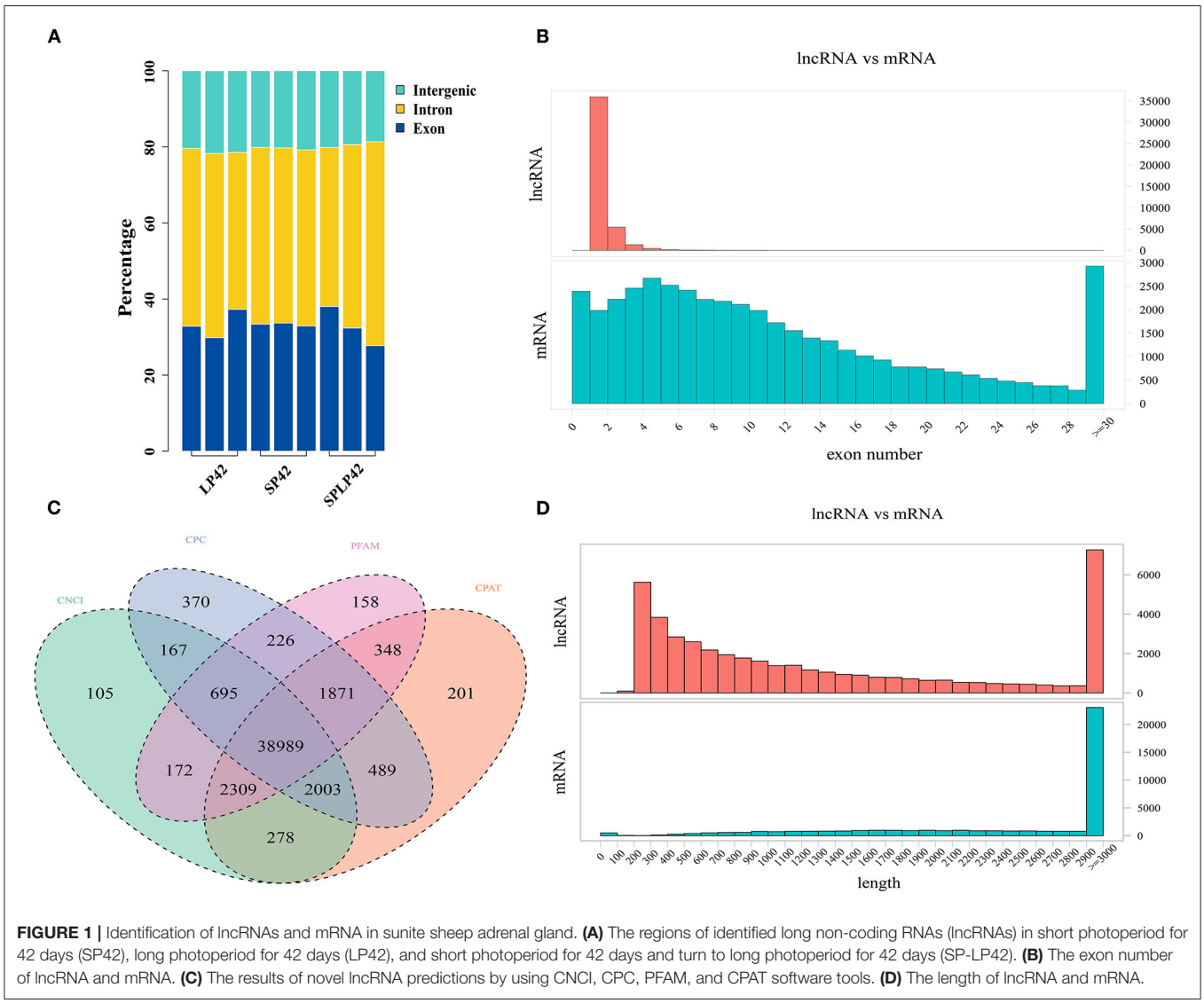


FIGURE 1 | Identification of lncRNAs and mRNA in sunite sheep adrenal gland. **(A)** The regions of identified long non-coding RNAs (lncRNAs) in short photoperiod for 42 days (SP42), long photoperiod for 42 days (LP42), and short photoperiod for 42 days and turn to long photoperiod for 42 days (SP-LP42). **(B)** The exon number of lncRNA and mRNA. **(C)** The results of novel lncRNA predictions by using CNCI, CPC, PFAM, and CPAT software tools. **(D)** The length of lncRNA and mRNA.

RESULTS

Identification of lncRNAs and mRNAs in the Adrenal Gland Tissue

The RNA-Seq raw data obtained in this study were subjected to quality control. The results are shown in **Table 1** and **Supplementary Table 1**. In total, SP42 ($n = 117{,}038{,}475$), LP42 ($n = 108{,}176{,}645$) and SP-LP42 ($n = 117{,}740{,}833$) clean reads of average were obtained, respectively, from adrenal gland tissues. Q30 base rate as the filtered data standards, the results show that the percentage of each sample more than 93.70%, above suggests that higher credibility. In comparison with the reference genome (Oar_v4.0) of *Ovis aries*, the mapping rate of each sample is >94%, which is a satisfactory sequencing results. Subsequently, regions in the genome with the identified lncRNAs were predicted (**Figure 1A**). We found that many of the lncRNAs belong to intron regions, followed by exon and intergenic regions (**Supplementary Table 1**). In addition, many of the lncRNAs were longer than 200 bp, with many in the range of about 2,900–3,000 bp in length, and the majority of lncRNAs have only two exons. Compared with lncRNAs, mRNAs have more than two exons on average, and most of the lengths are concentrated in the range of 2,900–3,000 bp (**Figures 1B,D**). We also identified novel lncRNAs by using CNCI, CPC, PFAM and CPAT software to predict the screened non-coding RNA. The results reveal that 38,989 novel lncRNAs were identified and that 29,695 novel lncRNAs were expressed in our samples, including lncRNAs ($n = 10{,}362$), antisense lncRNAs ($n = 2{,}462$) and intronic lncRNAs ($n = 16{,}871$) (**Figure 1C**; **Supplementary Table 2**).

Differential Expression Analysis of lncRNAs and mRNAs

Pursuant to the expression of differentially expressed lncRNA and mRNA (DELs, DEMs) in each sample, $|\log2\text{Ratio}| \geq 1$ and $q < 0.05$ as cut-off, we found 304, 1,110 and 928 DELs in LP42 vs. SP-LP42, SP42 vs. SP-LP42 and SP42 vs. LP42, respectively. The number of up-regulated genes was 120, 333 and 332, respectively, and the number of down-regulated genes was 184, 777 and 596 respectively. We also identified 144 DEMs (up-regulated 45, down-regulated 99) in LP42 vs. SP-LP42, 454 DEMs (up-regulated 74, down-regulated 380) in SP42 vs. LP42, and 506 DEMs (up-regulated 147, down-regulated 359) in SP42 vs. SP-LP42 (**Figure 2**; **Supplementary Table 3**). According to a base logarithm of 2 of expression about DEMs and DELs in each sample and the Euclidean distance was calculated, and then the overall clustering results of the samples were obtained by systematic clustering method (Hierarchical Cluster; **Figure 3**). An interesting phenomenon about the pattern of DELs is that cluster analysis showed SP-LP42A1 and LP42A2 as mixed groups, and the pattern of DEMs showed perfect groups which is divided into three parts (**Figures 3A,B**). As we expected, there were significant differences in DELs and DEMs between SP42 treated group and LP42 treated group. However, the expression pattern of DELs indicates that there may be a similar pattern between SP-LP42 and LP42, but the reason is not clear. This, therefore, led to subsequent mining key candidate lncRNA and mRNA transcripts mainly concentrated in the SP42 vs. LP42 comparison group.

Validation of RNA Sequencing Using RT-qPCR

To verify the sequencing reliability, seven lncRNAs (**Table 2**) were randomly selected from the three comparison groups and subjected to RT-qPCR testing. The relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The results found similar expression patterns using RNA-Seq and RT-qPCR (**Figure 4**).

Gene Enrichment Analysis

GO annotation and KEGG enrichment analysis were conducted using the identified target genes of DELs. Many GO terms related to ATP binding, Golgi organization, ATP-dependent helicase activity and ATPase activity (**Figure 5**; **Supplementary Table 4**). However, the KEGG pathway enrichment of the LP42 vs. SPLP42 group was not as significant as that of the other two groups.

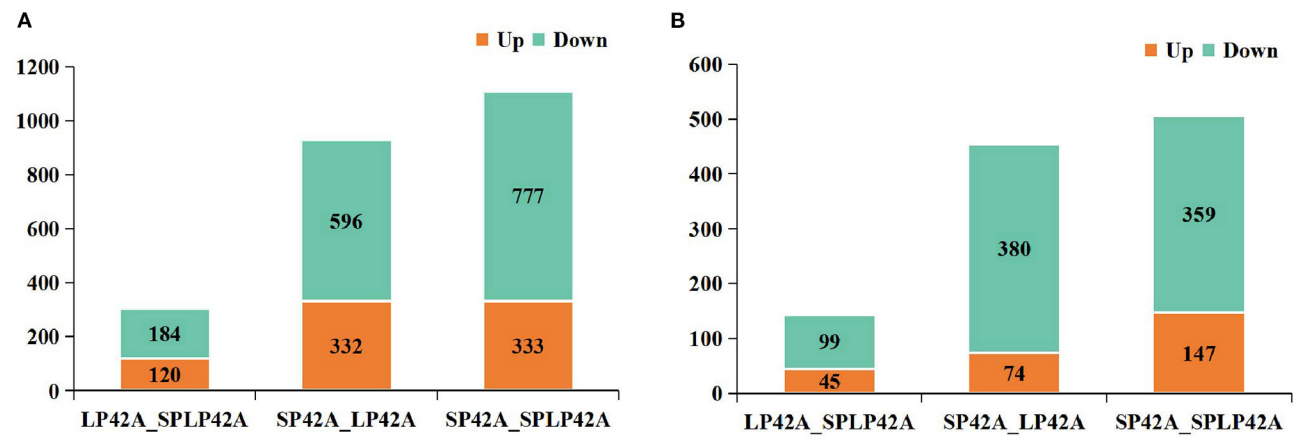


FIGURE 2 | The histogram of DELs and DEMs in different comparable groups. **(A)** DELs **(B)** DEMs. Where red and green represent up- or down regulation, respectively.

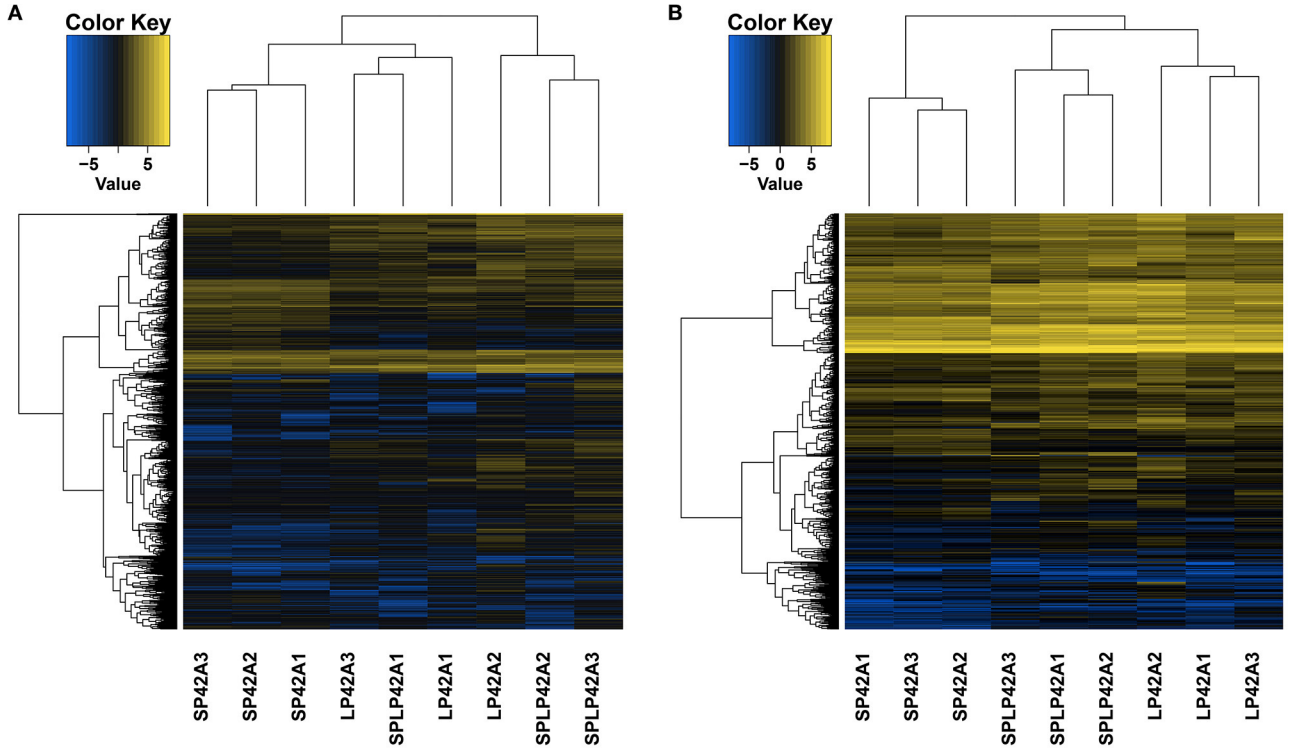


FIGURE 3 | The differentially expressed lncRNAs and mRNAs in each sample using Heat maps. According to a base logarithm of 2 of expression about DEMs and DELs in each sample and the Euclidean distance was calculated, and then the overall clustering results of the samples were obtained by systematic clustering method (Hierarchical Cluster), **(A)** lncRNAs, **(B)** mRNAs.

TABLE 2 | Real-time quantitative polymerase chain reaction primers and sizes of the amplification products of the selected lncRNAs and housekeeping genes.

lncRNA	Primer sequence (5′-3′)	Product size (bp)
LOC106991530/ XR_003591261.1	CTCCGGGAAACTTGGTCTCT	89
	CCCAGTTCTGCCAGGAGTTA	
LOC105609559	GAAGCCACCCAAATCCAGAC	182
	GAGCACCAACCATCCTCTCT	
LOC105609997	TGGCTTCCATGGACTGATGT	174
	AATCCACACTCCTCCCTTGG	
MSTRG.21610	GCGGAGGAAGTAGGCTCTAG	167
	CTCGCATCCAAAGCTCAGAC	
MSTRG.4985	AGGAAATCAACAACGGTGCC	80
	TTCCACGTTTCCTCTCCCTC	
MSTRG.21588	GCTATAGGAAGGGCTCTGGG	97
	GGCACAACTGAAGCAACTGA	
MSTRG.196373	CCGTGAACTTGGTGGCATAG	191
	TTCTCCTACCTGCCTCCTCT	
β-actin	CCAACCGTGAGAAGATGACC	97
	CCCGAGGCGT ACAGGGACAG	

The SP42 vs. SPLP42 group, and the SP42 vs. LP42 group, were shown to have similar pathways. Pathways associated with these two groups were related to TNF signaling, sphingolipid signaling, cancer, MAPK signaling, Hippo signaling and dopaminergic synapse (**Figure 6**; **Supplementary Table 5**).

Building lncRNA–mRNA Interaction Networks

To further describe the interaction between lncRNA, and its target genes, we constructed an interaction network of differentially expressed genes in the SP42 vs. LP42 comparison group. A lncRNA/mRNA co-expression network was constructed using 82 differentially expressed lncRNAs and 11 target genes involved in reproductive-related pathways (**Figure 7**). Twenty up-regulated lncRNA, and 60 down-regulated lncRNA, were identified. Of these, only 2 of the 60 down-regulated RNA are known lncRNAs. The remainder is novel lncRNA. *FGF16* is the only up-regulated gene (of the 11 target genes), and the rest of the target genes are down-regulated. Interestingly, only two lncRNA have a cis-regulatory relationship with their target genes. The remainder is trans-regulatory relationships.

DISCUSSION

The influence of long non-coding RNA on the reproductive function of sheep had been extensively investigated. Several genes that affect sheep reproduction had been found in

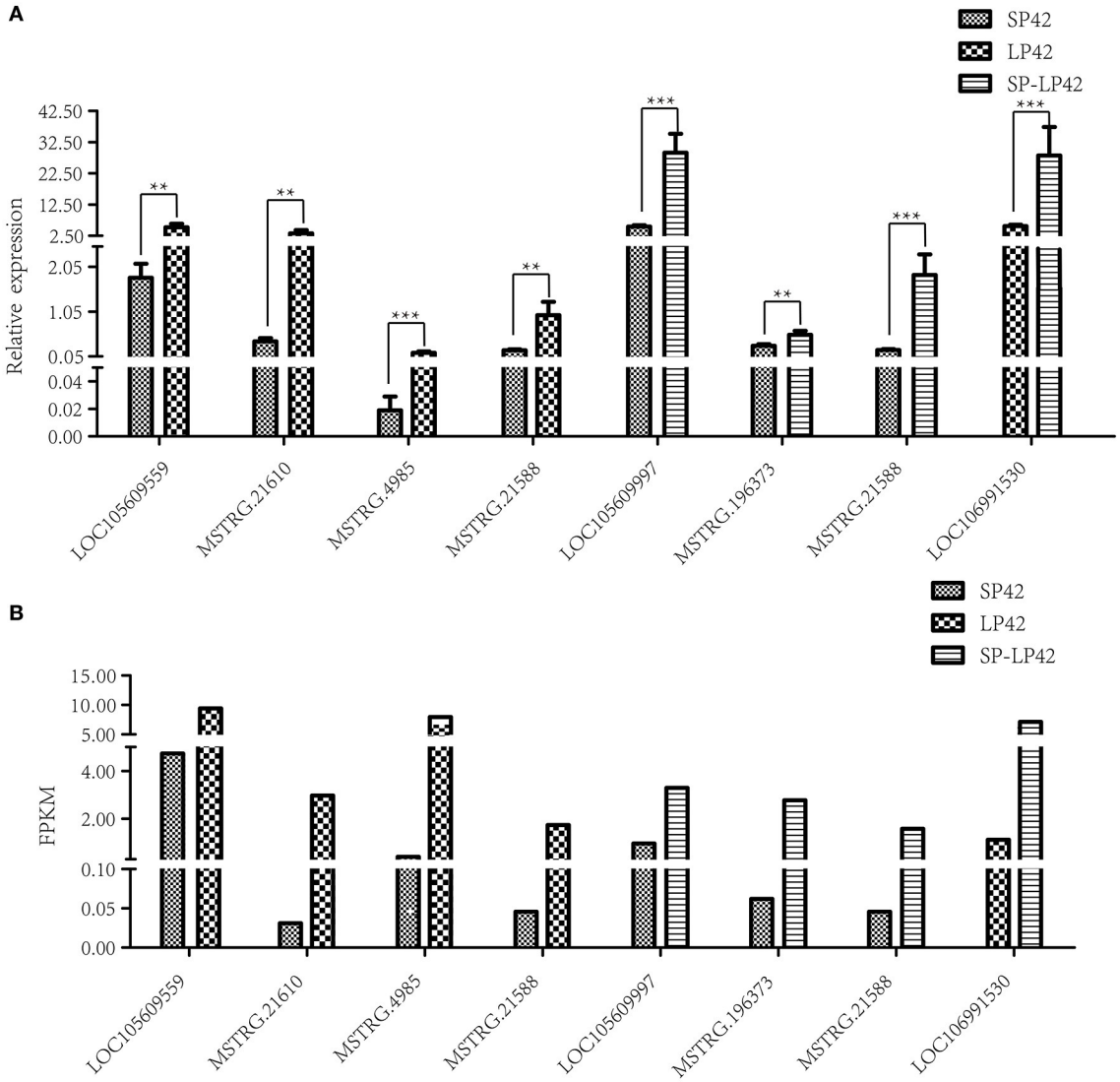


FIGURE 4 | Validation of RNA-Sequencing (RNA-Seq) data using reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR). Different types of rectangles represent different light period processing. The figure ** and *** represents the p value $\leq$ 0.05 and 0.01, respectively. **(A)** RT-qPCR, **(B)** RNA-seq.

lncRNA studies of the hypothalamus (29) and adrenal tissue (33). The hypothalamus was the center that regulates the life activities of mammals, including survival, growth and development, and reproduction. Our team had conducted in-depth research on the hypothalamus and ovaries (46). It was well-known that seasonal estrus is the key factor affecting sheep reproduction, but the change of photoperiod was the key factor affecting the change of seasonal estrus rhythm (9, 10). Photoperiod could be considered as a source of exogenous stress in animals, and the adrenal gland is a key organ to deal with stress response. The OVX + E_2 model was a good model to study the effect of light period on reproduction (21, 22, 47), Therefore, we used this model for transcriptome sequencing analysis.

We detected a large number of lncRNA and mRNA in the adrenal gland of Sunite sheep by RNA-Seq and counted the length, and the number of exons. We found that the length of lncRNA was less than that of mRNA. Some studies had shown that the length, and exon number, of lncRNA in the sheep hypothalamus was larger than that of the goat hypothalamus (26, 48). Further studies had found that the lncRNA length of the sheep hypothalamus was also longer than that of mice, however, the number of exons was less than that of mice (29, 49). Our study found that the length, and exon number, of lncRNA and mRNA, in sheep adrenal tissue was different from sheep hypothalamus tissue. In particular, the regions of identified lncRNA were significantly different from that of other sheep hypothalamus, more lncRNA were clustered in an intron,

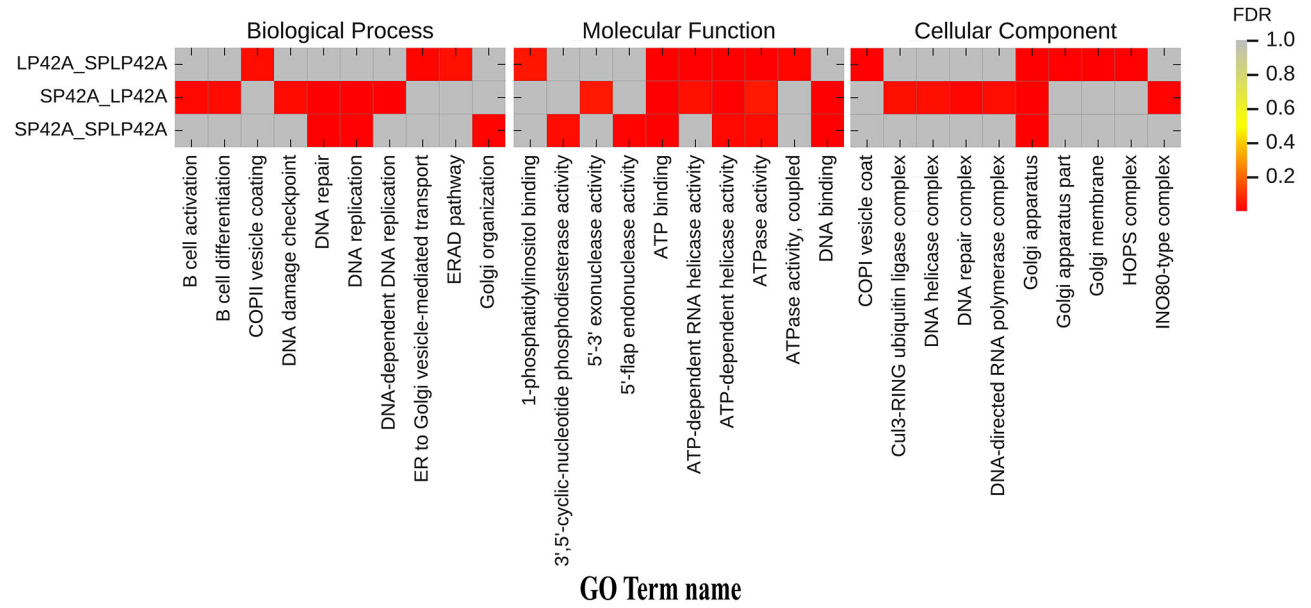


FIGURE 5 | Histogram of GO enrichment of target gene of DELs. Heat maps showing the GO items enriched in the three comparison groups of LP42 vs. SPLP42, SP42 vs. LP42 and SP42 vs. SPLP42.

followed by exon and intergenic and the type of intronic RNA accounts for more than 56.8% (46). Therefore, lncRNAs were tissue- and species-specific (50).

Cluster analysis of lncRNAs and mRNAs showed that the three samples of the SP42-lncRNA group and all samples of mRNAs were perfectly clustered together, respectively, but the lncRNA of LP42 group and the SP-LP42 group were not completely classified. Thus, did it mean that the similarity of lncRNA expression between the LP42 group and the SP-LP42 group? Is it because the variation and restoration of photoperiod mode also leads to the change and restoration of lncRNA expression pattern? Is SP42 vs. SP-LP42 consistent with the differentially expressed genes of SP42 vs. LP42? To answer these questions, we carried out GO and KEGG enrichment analysis. We found our inference was correct, the SP42 vs. LP42 comparison group and the SP42 vs. SP-LP42 comparison group in the case of $q \leq 0.01$, GO term and KEGG pathway are the same. However, LP42 vs. SP-LP42 did not find KEGG pathway with $q \leq 0.01$. Therefore, we selected the pathway with significant enrichment of KEGG in the SP42 vs. LP42 comparison group to screen candidate genes affecting reproduction.

The fibroblast growth factor 16 (*FGF16*) gene was related to oocyte maturation. In dairy cows the expression of the *FGF16* gene was correlated to oocyte quality (51). In summer, when oocyte quality was low, the expression of *FGF16* was low. Conversely, in winter, when oocyte quality was high, the expression of *FGF16* was high. We found that the expression of *FGF16* gene was up-regulated in the comparative group of SP42 vs. LP42; indicating, that in adrenal tissue, the expression of *FGF16* gene in long photoperiod was nearly 10 times lower than that in short photoperiod. Finally, two important up-regulation of *FGF16* gene lncRNA (MSTRG.242136, MSTRG.236582), and 11 down-regulation of lncRNA were identified in our study. Among them, the expression abundance of MSTRG.176476 lncRNA was the highest, and |Log2 Fold Change| was about 2. In addition, although the expression abundance was not high, the largest |Log2 Fold Change| close to 7 is MSTRG.21610 lncRNA, but its effect on *FGF16* gene expression remains to be further verified.

Placental growth factor (*PLGF*) was another gene that may be associated with reproductive function, most notably with embryo implantation (52, 53). *PLGF* was a member of the vascular endothelial growth factor family of proangiogenic factors regulated angiogenesis and microvessel density (MVD) (54, 55). Moreover, the serum level of *PLGF* had been positively correlated with fecundity in Hu sheep (56). The differential expression of *PLGF* gene in the SP42 vs. LP42 comparison group was also found in our study, and |Log2 Fold Change| was close to 2.4. As we know that the retina was the window for receiving light signals and was filled with microvessels. Study had shown that *PLGF* was related to retinal angiogenesis (57). Therefore whether *PLGF* gene affected light signal reception through related pathways, and thus, affected seasonal estrus needs, need to be further explored. Interestingly, the differentially expressed gene cyclin-dependent kinase inhibitor 1A (*CDKN1A*), which was found in our study and |Log2 Fold Change| was close to 1.6, $p \leq 0.01$, played an important role in the apoptosis of vitreous microvascular epithelial cells (58). At the same time, another key candidate gene Endothelial differentiation gene1 (*EDG1*) for angiogenesis was also found in our study and |Log2 Fold Change| was close to 1.4, $p \leq 0.01$. Specifically, *EDG1* also known as sphingosine 1–phosphate receptor 1 (*S1PR1*) which belong to the rhodopsin family, was involved in angiogenesis. This family had been considered to

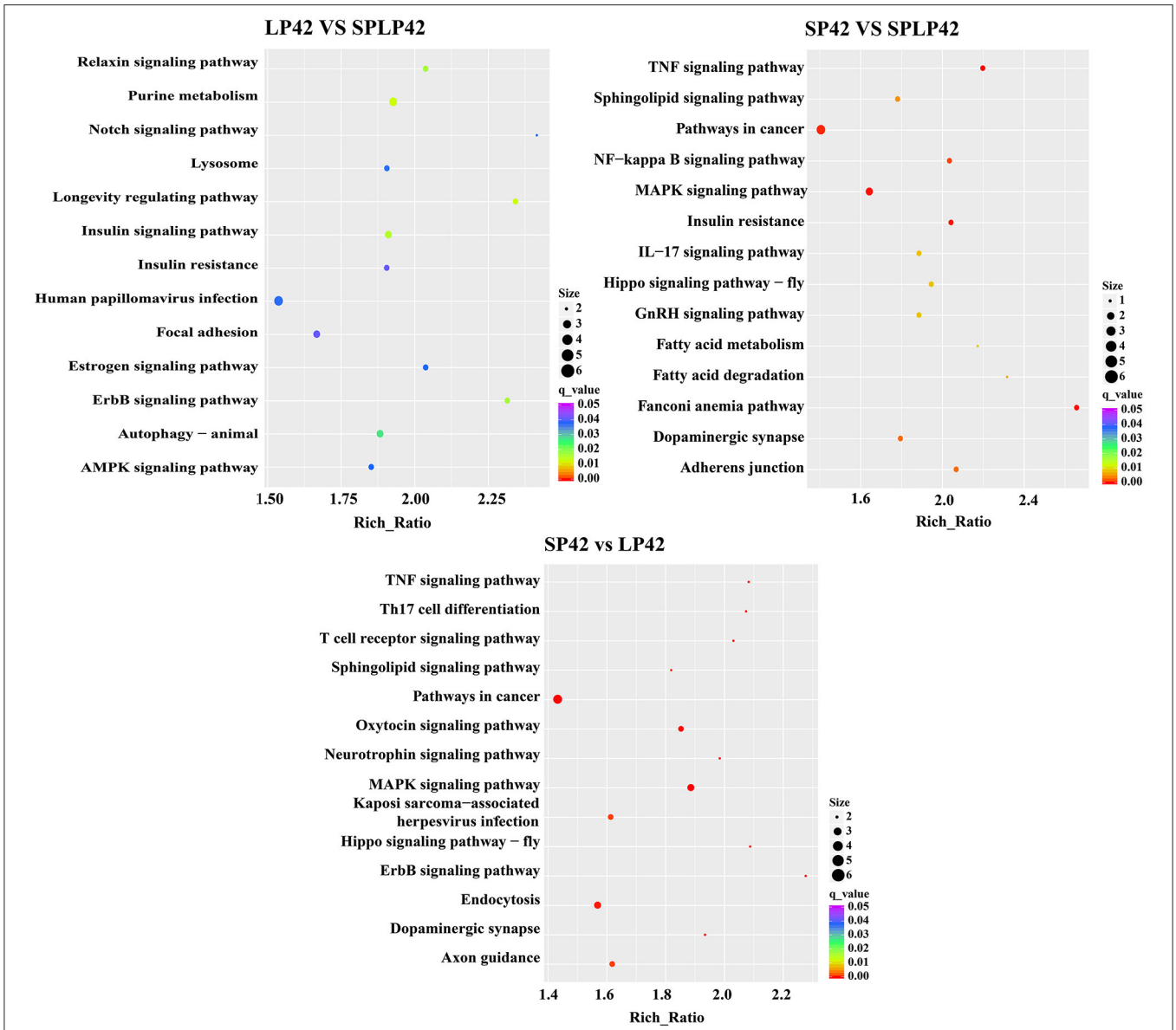


FIGURE 6 | KEGG pathway enrichment analysis of target gene of DELs. The significant enriched KEGG pathways in the three comparison groups of LP42 vs. SPLP42, SP42 vs. LP42 and SP42 vs. SPLP42.

be typical members of the rhodopsin superfamily. The function of most opsins was split into two steps: light absorption and G-protein activation. In addition, *EDG1* expression had been observed in ovarian tissues and the family of *S1PR1* also had been reported to play an important role in ovarian angiogenesis, suggesting that the *EDG1* signal may regulate ovarian angiogenesis. Generally, ovarian angiogenesis seems to be one of the factors responsible for follicular development. Consequently, *EDG1* was currently used as a genetic marker for reproductive traits in cattle because there was a significant correlation between *EDG1* polymorphism and the age of first birth in cattle (59).

In our experimental design, photoperiod as a unique variable and the only exogenous stress, the experimental samples only through the retina to receive light stimulation to change the biological clock to further change their hormone secretion. Then whether it is possible to explore the secretion of seasonal estrous hormones according to the mechanism of the retina may be a valuable topic. It was well-known that GnRH played an important role in sheep reproduction. Among the differential genes identified, semaphorin 7 (*SEMA7A*), |Log2 Fold Change| was close to 1.4, was reported to be closely related to the development of mouse GnRH-1 neurons system (60). Among the differential genes identified by us, the gene with the

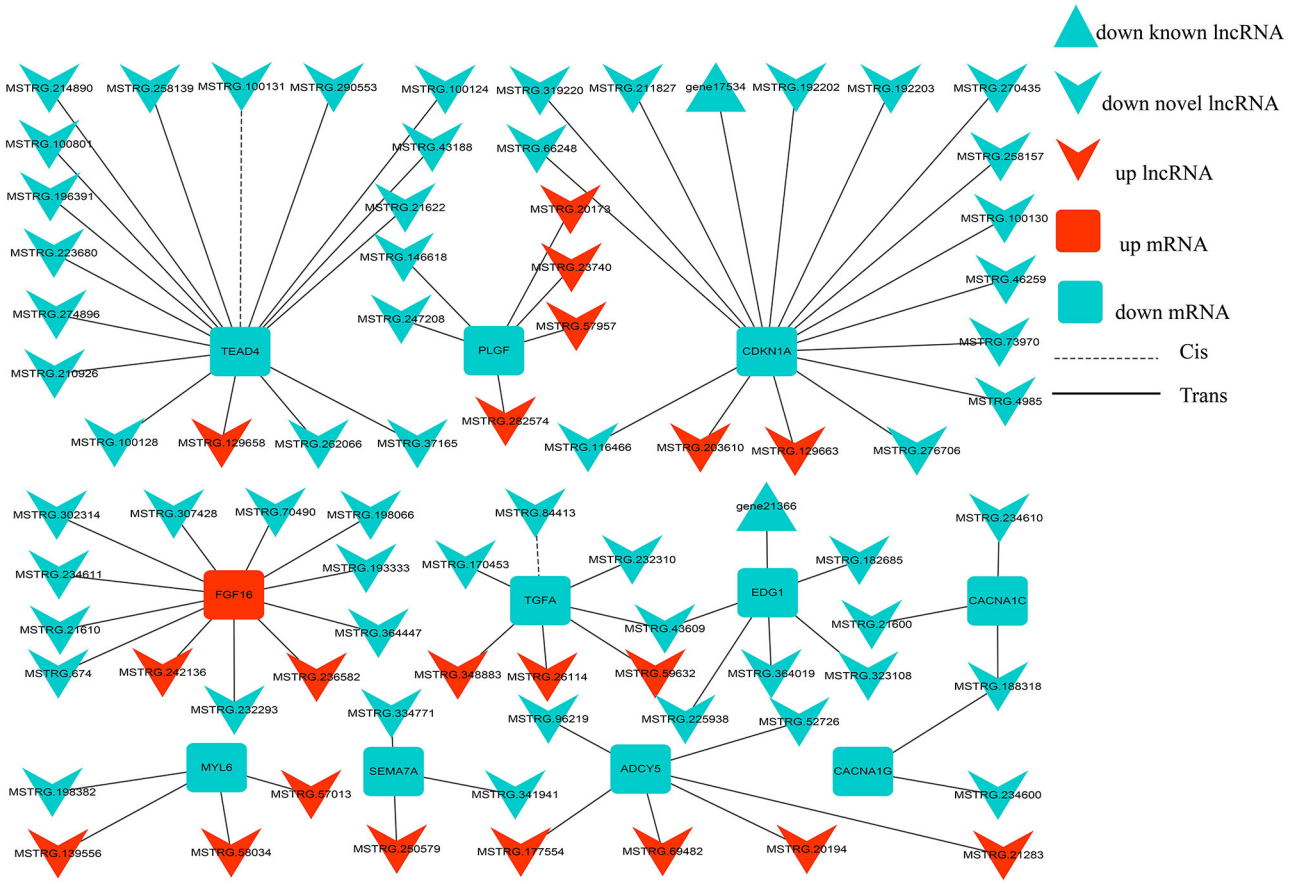


FIGURE 7 | lncRNA-mRNA interaction networks. The interaction networks of lncRNAs and their corresponding target genes in SP42 vs. LP42, where the solid and dashed lines represent trans- and cis-regulation functions, respectively; red and green represent up- and downregulation, respectively; and round rectangle, V and triangle represent mRNAs, novel lncRNAs and known lncRNA, respectively.

similar function was voltage-dependent calcium channel L type alpha-1C (*CACNA1C*), which had been proved to be a key candidate for precocious puberty in Jining gray goats (61). Interestingly, the gene had also been shown to be closely related to GnRH. It co-ordinatively participated in ERK activation and caused the increase of FSH and LH secretion in the GnRH signal pathway (62). In addition, adenylate cyclase 5 (*ADCY5*) which |Log2 Fold Change| was close to 2 in our study had also been proved to be a key candidate gene affecting the fecundity of dairy cows (63). Thus, whether the gene affects the reproductive ability of sheep through the adrenal gland under different light conditions needs to be further explored.

CONCLUSION

In conclusion, this study provided lncRNA and mRNA expression profiling in the adrenal gland of sheep during different photoperiods. Several photoperiod-induced targeting key genes of seasonal reproduction (*FGF16*, *PLGF*, *CDKN1A*, *SEMA7A*, *EDG1*, *CACNA1C* and *ADCY5*) were predicted in the adrenal gland of sheep. These results may provide a solid molecular basis for follow-up studies on seasonal estrus in sheep.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The animal study was reviewed and approved by the Science Research Department (in charge of animal welfare issues) of the Institute of Animal Science, Chinese Academy of Agricultural Sciences (IAS-CAAS) (Beijing, China) and ethical approval was given by the Animal Ethics Committee of the IAS-CAAS (No. IAS2018-3). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

QiuL and MC designed the research. XD wrote the paper and performed the study. RD, XD, and QinL collected the data. XD and XH analyzed data. MC revised the final manuscript. All authors reviewed the manuscript and approved the final version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (32172704), China Agriculture Research System of MOF and MARA (CARS-38), the Agricultural Science and Technology Innovation Program of China (CAAS-ZDRW202106 and ASTIP-IAS13).

ACKNOWLEDGMENTS

We are grateful to the Xiaosheng Zhang, Jinlong Zhang and Xiaofei Guo in Tianjin Institute of Animal Sciences for the ovariectomy, estrogen implantation, and ewes feeding. Thanks to Xiaohan Cao, Xiaoyu Li, Xinlong Dong and Qing Xia for their help in sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.816241/full#supplementary-material

Supplementary Table 1 | The basic information of lncRNA sequencing data.

Supplementary Table 2 | The FPKM values of both known and novel lncRNAs in SP42, LP42, and SP-LP42.

Supplementary Table 3 | The information of lncRNAs and mRNA expressed in SP42 vs. LP42, SP42 vs. SP42-LP42 and LP42 vs. SP42-LP42.

Supplementary Table 4 | GO enrichment annotation for target genes of lncRNAs at the MF, BP, and CC level in SP42 vs. LP42, SP42 vs. SP42-LP42 and LP42 vs. SP42-LP42.

Supplementary Table 5 | KEGG enrichment pathway for target genes of lncRNAs in SP42 vs. LP42, SP42 vs. SP42-LP42 and LP42 vs. SP42-LP42.

REFERENCES

1. Lincoln GA, Short RV. Seasonal breeding: nature's contraceptive. *Recent Prog Horm Res.* (1980) 36:1–43. doi: 10.1016/B978-0-12-571136-4.50007-3
2. Rani S, Kumar V. Photoperiodic regulation of seasonal reproduction in higher vertebrates. *Indian J Exp Biol.* (2014) 52:413—9. doi: 10.1016/j.freeradbiomed.2014.01.023
3. Yu Q. Biological clock: the oscillator of gene expression. *Sci China Life Sci.* (2018) 61:128–30. doi: 10.1007/s11427-017-9239-6
4. Dahl GE Evans NP, Moenter SM, Karsch FJ. The thyroid gland is required for reproductive neuroendocrine responses to photoperiod in the ewe. *Endocrinology.* (1994) 135:10–5. doi: 10.1210/endo.135.1.8013340
5. Thiery JC, Malpaux B. Seasonal regulation of reproductive activity in sheep: modulation of access of sex steroids to the brain. *Ann N Y Acad Sci.* (2003) 1007:169–75. doi: 10.1196/annals.1286.017
6. He XY Di R, Hu WP, Wang XY, Liu QY, Chu MX. Research progress on neuroendocrinology controlling sheep seasonal reproduction. *Acta Veterinaria et Zootechnica Sinica.* (2018) 49:18–25. doi: 10.11843/j.issn.0366-6964.2018.01.003
7. Batailler M, Chesneau D, Derouet L, Butruille L, Segura S, Migaud M, et al. Pineal-dependent increase of hypothalamic neurogenesis contributes to the timing of seasonal reproduction in sheep. *Sci Rep.* (2018) 8:6188. doi: 10.1038/s41598-018-24381-4
8. Karsch FJ, Bittman EL, Foster DL, Goodman RL, Legan SJ, Robinson JE. Neuroendocrine basis of seasonal reproduction. *Recent Prog Horm Res.* (1984) 40:185–225. doi: 10.1016/B978-0-12-571140-1.50010-4
9. Weems PW, Goodman RL, Lehman MN. Neural mechanisms controlling seasonal reproduction: principles derived from the sheep model and its comparison with hamsters. *Front Neuroendocrinol.* (2015) 37:43–51. doi: 10.1016/j.yfrne.2014.12.002
10. Masumoto KH, Ukai-Tadenuma M, Kasukawa T, Nagano M, Shigeyoshi Y, Ueda HR, et al. Acute induction of Eya3 by late-night light stimulation triggers TSH beta expression in photoperiodism. *Curr Biol.* (2010) 20:2199–206. doi: 10.1016/j.cub.2010.11.038
11. Hanon EA, Lincoln GA, Fustin JM, Dardente H, Masson-Pevet M, Hazlerigg DG, et al. Ancestral TSH mechanism signals summer in a photoperiodic mammal. *Curr Biol.* (2008) 18:1147–52. doi: 10.1016/j.cub.2008.06.076
12. Acevedo-Rodriguez A, Kauffman AS, Cherrington BD. Emerging insights into hypothalamic-pituitary-gonadal axis regulation and interaction with stress signaling. *J Neuroendocrinol.* (2018) 30:e12590. doi: 10.1111/jne.12590
13. Handa RJ, Weiser MJ. Gonadal steroid hormones and the hypothalamic-pituitary-adrenal axis. *Front Neuroendocrinol.* (2014) 35:197–220. doi: 10.1016/j.yfrne.2013.11.001
14. Abdoli R, Zamani P, Mirhoseini SZ, Ghavi Hossein-Zadeh N, Nadri S. A review on prolificacy genes in sheep. *Reprod Domest Anim.* (2016) 51:631–7. doi: 10.1111/rda.12733
15. Ciechanowska M, Apot M, Paruszewska E, Radawiec W, Przekop F. The influence of dopaminergic system inhibition on biosynthesis of gonadotrophin-releasing hormone (GnRH) and GnRH receptor in anestrus sheep; hierarchical role of kisspeptin and RF amide-related peptide-3 (RFRP-3). *Reprod Fertil Dev.* (2018) 30:672–80. doi: 10.1071/RD16309
16. Luo E, Stephens SBZ, Chaing S, Munaganuru N, Kauffman AS, Breen KM. Corticosterone blocks ovarian cyclicity and the LH surge via decreased kisspeptin neuron activation in female mice. *Endocrinology.* (2016) 157:1187–99. doi: 10.1210/en.2015-1711
17. Dobson H, Routly JE, Smith RF. Understanding the trade-off between the environment and fertility in cows and ewes. *Anim Reprod.* (2020) 17:e20200017. doi: 10.1590/1984-3143-ar2020-0017
18. Chen K, Carey LC, Liu J, Valego NK, Tatter SB, Rose JC. The effect of hypothalamic-pituitary disconnection on the renin-angiotensin system in the late-gestation fetal sheep. *Am J Physiol Regul Integr Comp Physiol.* (2005) 288:R1279–1287. doi: 10.1152/ajpregu.00560.2004
19. Lincoln GA, Clarke IJ, Hut RA, Hazlerigg DG. Characterizing a mammalian circannual pacemaker. *Science.* (2006) 314:1941–4. doi: 10.1126/science.1132009
20. Lincoln GA. Neuroendocrine regulation of seasonal gonadotrophin and prolactin rhythms: lessons from the Soay ram model. *Reprod Suppl.* (2002) 59:131–47. doi: 10.1530/rep.0.1230067
21. Johnson ML, Grazul-Bilska AT, Reynolds LP, Redmer DA. Prion (PrPC) expression in ovine uteroplacental tissues increases after estrogen treatment of ovariectomized ewes and during early pregnancy. *Reproduction.* (2014) 148:1–10. doi: 10.1530/REP-13-0548
22. Weems P, Smith J, Clarke IJ, Coolen LM, Goodman RL, Lehman MN. Effects of season and estradiol on KNDy neuron peptides, colocalization with D2 dopamine receptors, and dopaminergic inputs in the ewe. *Endocrinology.* (2017) 158:831–41. doi: 10.1210/en.2016-1830
23. Derrien T, Johnson R, Bussotti G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* (2012) 22:1775–89. doi: 10.1101/gr.132159.111

24. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* (2018) 172:393–407. doi: 10.1016/j.cell.2018.01.011
25. Miao XY, Luo QM, Zhao HJ, Qin XY. Co-expression analysis and identification of fecundity-related long non-coding RNAs in sheep ovaries. *Sci Rep-UK.* (2016) 6:39398. doi: 10.1038/srep39398
26. Gao XX, Ye J, Yang C, Zhang KF, Cao HG, Ling YH, et al. Screening and evaluating of long noncoding RNAs in the puberty of goats. *BMC Genomics.* (2017) 18:14. doi: 10.1186/s12864-017-3578-9
27. Gao XX, Zhang YH, Ling YH, Huang WP, Zhang XR, Cao ZB, et al. RNA-seq analysis of lncRNA-controlled developmental gene expression during puberty in goat & rat. *BMC Genet.* (2018) 19:19. doi: 10.1186/s12863-018-0608-9
28. Zheng J, Wang ZB, Yang H, Yao XL, Wang F, Zhang YL, et al. Pituitary transcriptomic study reveals the differential regulation of lncRNAs and mRNAs related to prolificacy in different FecB genotyping sheep. *Genes-Basel.* (2019) 10:157. doi: 10.3390/genes10020157
29. Zhang ZB, Tang JS Di R, Zhang JL, Hu WP, Chu MX, et al. Comparative transcriptomics reveal key sheep (*Ovis aries*) hypothalamus lncRNAs that affect reproduction. *Animals-Basel.* (2019) 9:152. doi: 10.3390/ani9040152
30. Gao Y, Chen F, Kong QQ, Ning SF, Luo MJ, Tan JH, et al. Stresses on female mice impair oocyte developmental potential: effects of stress severity and duration on oocytes at the growing follicle stage. *Reprod Sci.* (2016) 23:1148–57. doi: 10.1177/1933719116630416
31. Yuan HJ, Han X, He N, Wang GL, Gao M, Tan JH, et al. Glucocorticoids impair oocyte developmental potential by triggering apoptosis of ovarian cells via activating the Fas system. *Sci Rep-UK.* (2016) 6:24036. doi: 10.1038/srep24036
32. Whirledge S, Cidlowski JA. Glucocorticoids and reproduction: traffic control on the road to reproduction. *Trends Endocrin Met.* (2017) 28:399–415. doi: 10.1016/j.tem.2017.02.005
33. Xia Q, Li QL, Gan SQ, Guo XF, Zhang JL, Chu MX, et al. Exploring the roles of fecundity-related long non-coding RNAs and mRNAs in the adrenal glands of small-tailed Han Sheep. *BMC Genet.* (2020) 21:39. doi: 10.1186/s12863-020-00850-6
34. La YF, He XY, Zhang XS, Zhang JL, Hu WP, Chu MX, et al. Comprehensive analysis of differentially expressed profiles of mRNA, lncRNA, and circRNA in the uterus of seasonal reproduction sheep. *Genes-Basel.* (2020) 11:301. doi: 10.3390/genes11030301
35. He XY, Tao L, Zhong YJ Di R, Xia Q, Chu MX, et al. Photoperiod induced the pituitary differential regulation of lncRNAs and mRNAs related to reproduction in sheep. *PeerJ.* (2021) 9:e10953. doi: 10.7717/peerj.10953
36. Xia Q, Chu MX, He XY, Zhang XS, Guo XF, Di R, et al. Identification of photoperiod-induced lncRNAs and mRNAs in pituitary pars tuberalis of sheep. *Front Vet Sci.* (2021) 8:644474. doi: 10.3389/fvets.2021.644474
37. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* (2016) 11:1650–67. doi: 10.1038/nprot.2016.095
38. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* (2015) 33:290. doi: 10.1038/nbt.3122
39. Sun L, Luo HT, Bu DC, Liu YN, Chen RS, Zhao Y, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* (2013) 41:e166. doi: 10.1093/nar/gkt646
40. Kong L, Zhang Y, Ye ZQ, Zhao SQ, Wei L, Gao G, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* (2007) 35:W345–9. doi: 10.1093/nar/gkm391
41. El-Gebali S, Mistry J, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* (2019) 47:D427–32. doi: 10.1093/nar/gky995
42. Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP Li W. CPAT Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* (2013) 41:e74. doi: 10.1093/nar/gkt006
43. Trapnell C, Williams BA, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* (2010) 28:511–5. doi: 10.1038/nbt.1621
44. Wang LK, Feng ZX, Wang X, Wang XW, Zhang XG. DEMseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* (2010) 26:136–8. doi: 10.1093/bioinformatics/btp612
45. Kolde R, Kolde MR. Package 'Pheatmap'. *R Package.* (2015) 1:790.
46. Chen S, Guo XF, He XY, Di R, Wang XY, Chu MX, et al. Transcriptome analysis reveals differentially expressed genes and long non-coding RNAs associated with fecundity in sheep hypothalamus with different FecB genotypes. *Front Cell Dev Biol.* (2021) 9:633747. doi: 10.3389/fcell.2021.633747
47. Merkley CM, Coolen LM, Goodman RL, Lehman MN. Evidence for changes in numbers of synaptic inputs onto KNDy and GnRH neurones during the preovulatory LH surge in the ewe. *J Neuroendocrinol.* (2015) 27:624–35. doi: 10.1111/jne.12293
48. Flegel C, Manteniotis S, Osthold S, Hatt H, Gisselmann G. Expression profile of ectopic olfactory receptors determined by deep sequencing. *PLoS ONE.* (2013) 8:e55368. doi: 10.1371/journal.pone.0055368
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303
50. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Rinn JL, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* (2011) 25:1915–27. doi: 10.1101/gad.17446611
51. Ferreira RM, Chiaratti MR, Macabelli CH, Smith LC, Meirelles FV, Baruselli PS, et al. The infertility of repeat-breeder cows during summer is associated with decreased mitochondrial DNA and increased expression of mitochondrial and apoptotic genes in oocytes. *Biol Reprod.* (2016) 94:66. doi: 10.1095/biolreprod.115.133017
52. Nejabati HR, Latifi Z, Ghasemnejad T, Fattahi A, Nouri M. Placental growth factor (*PLGF*) as an angiogenic/inflammatory switcher: lesson from early pregnancy losses. *Gynecol Endocrinol.* (2017) 33:668–74. doi: 10.1080/09513590.2017.1318375
53. Binder NK, Evans J, Salamonsen LA, Gardner DK. Kaitu'u-Lino TJ, Hannan NJ. Placental growth factor is secreted by the human endometrium and has potential important functions during embryo development and implantation. *PLoS ONE.* (2016) 11:e163096. doi: 10.1371/journal.pone.0163096
54. Chen CN, Hsieh FJ, Cheng YM, Su YN, Chang KJ, Lee PH, et al. The significance of placenta growth factor in angiogenesis and clinical outcome of human gastric cancer. *Cancer Lett.* (2004) 213:73–82. doi: 10.1016/j.canlet.2004.05.020
55. Odorisio T, Schietroma C, Tiveron C, Tatangelo L, Failla CM, Zambruno G, et al. Mice overexpressing placenta growth factor exhibit increased vascularization and vessel permeability. *J Cell Sci.* (2002) 115:2559–67. doi: 10.1242/jcs.115.12.2559
56. Gao XX, Li XD, An SY, Wei ZY, Zhang GM, Wang F, et al. Long non-coding RNA366.2 controls endometrial epithelial cell proliferation and migration by upregulating WNT6 as a ceRNA of miR-1576 in sheep uterus. *Biochim Biophys Acta Gene Regul Mech.* (2020) 1863:194606. doi: 10.1016/j.bbagrm.2020.194606
57. Van Bergen T, Etienne I, Cunningham F, Moons L, Feyen JHM, Stitt AW, et al. The role of placental growth factor (*PLGF*) and its receptor system in retinal vascular diseases. *Prog Retin Eye Res.* (2019) 69:116–36. doi: 10.1016/j.preteyeres.2018.10.006
58. Nayak G, Odaka Y, Molkentin JD, Trumpp A, Williams B, Rao S, et al. Developmental vascular regression is regulated by a Wnt/beta-catenin, *MYC* and *CDKN1A* pathway that controls cell proliferation and cell death. *Development.* (2018) 145:dev154898. doi: 10.1242/dev.154898
59. Kawaguchi F, Tsuchimura M, Oyama K, Maruyama S, Mannen H, Sasazaki S, et al. Effect of DNA markers on the fertility traits of Japanese Black cattle for improving beef quantity and quality. *Arch Anim Breed.* (2020) 63:9–17. doi: 10.5194/aab-63-9-2020
60. Messina A, Ferraris N, Kramer PR, Derijck AA, Adolfs Y, Fasolo A, et al. Dysregulation of Semaphorin7A/beta1-integrin signaling leads to defective GnRH-1 cell migration, abnormal gonadal development and altered fertility. *Hum Mol Genet.* (2011) 20:4759–74. doi: 10.1093/hmg/ddr403
61. Su F, Guo XL, Wang YC, Wang YD, Cao GL, Jiang YL. Genome-wide analysis on the landscape of transcriptomes and their relationship

with DNA methylomes in the hypothalamus reveals genes related to sexual precocity in Jining Gray Goats. *Front Endocrinol.* (2018) 9:501. doi: 10.3389/fendo.2018.00501
62. Dang AK, Murtazina DA, Magee C, Navratil AM, Clay CM, Amberg GC. GnRH evokes localized subplasmalemmal calcium signaling in gonadotropes. *Mol Endocrinol.* (2014) 28:2049–59. doi: 10.1210/me.2014-1208
63. Cai ZX, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics.* (2019) 20:29. doi: 10.1186/s12864-019-5638-9

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

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

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fvets.2022.842074

Generation of Heritable Prominent Double Muscle Buttock Rabbits via Novel Site Editing of Myostatin Gene Using CRISPR/Cas9 System

Yalin Zheng[1†], Yu Zhang[1†], Liyan Wu[1], Hasan Riaz[3], Zhipeng Li[1], Deshun Shi[1], Saif ur Rehman[1], Qingyou Liu[1,2*] and Kuiqing Cui[1*]*

[1] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning, China, [2] Guangdong Provincial Key Laboratory of Animal Molecular Design and Precise Breeding, School of Life Science and Engineering, Foshan University, Foshan, China, [3] Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan

OPEN ACCESS

Edited by:
Marco Milanesi,
University of Tuscia, Italy

Reviewed by:
Emiliano Lasagna,
University of Perugia, Italy
Shi-Yi Chen,
Sichuan Agricultural University, China

***Correspondence:**
Saif ur Rehman
saif_ali28@yahoo.com
Qingyou Liu
qyliu-gene@gxu.edu.cn
Kuiqing Cui
kqcui@126.com

[†] These authors have contributed equally to this work

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

Received: 23 December 2021
Accepted: 18 March 2022
Published: 20 May 2022

Citation:
Zheng Y, Zhang Y, Wu L, Riaz H, Li Z, Shi D, Rehman Su, Liu Q and Cui K (2022) Generation of Heritable Prominent Double Muscle Buttock Rabbits via Novel Site Editing of Myostatin Gene Using CRISPR/Cas9 System. Front. Vet. Sci. 9:842074. doi: 10.3389/fvets.2022.842074

Rabbits have been domesticated for meat, wool, and fur production, and have also been cherished as a companion, artistic inspiration, and an experimental model to study many human diseases. In the present study, the muscle mass negative regulator gene myostatin (*MSTN*) was knocked out in rabbits at two novel sites in exon3, and the function of these mutations was determined in subsequent generations. The prominent double muscle phenotype with hyperplasia or hypertrophy of muscle fiber was observed in the *MSTN*-KO rabbits, and a similar phenotype was confirmed in the F1 generation. Moreover, the average weight of 80-day-old *MSTN*-KO rabbits (2,452 ± 63 g) was higher than that of wild-type rabbits (2,393.2 ± 106.88 g), and also the bodyweight of *MSTN*-KO rabbits (3,708 ± 43.06g) was significantly higher ($P < 0.001$) at the age of 180 days than wild-type (WT) rabbits (3,224 ± 48.64g). In *MSTN*-KO rabbits, fourteen rabbit pups from the F1 generation and thirteen from the F2 generation stably inherited the induced *MSTN* gene mutations. Totally, 194 pups were produced in the F1 generation of which 49 were *MSTN*-KO rabbits, while 47 pups were produced in the F2 generation of which 20 were edited rabbits, and the ratio of edited to wild-type rabbits in the F2 generation was approximately 1:1. Thus, we successfully generated a heritable double muscle buttocks rabbits via myostatin mutation with CRISPR/Cas9 system, which could be valuable in rabbit's meat production and also a useful animal model to study the development of muscles among livestock species and improve their important economic traits as well as the human muscle development-related diseases.

Keywords: rabbits, Cas9, *MSTN* gene, knock-out, double muscle buttocks rabbits

INTRODUCTION

Rabbits are long-eared ground-dwelling mammals belonging to the family *Leporidae* order *Lagomorpha* and domesticated about 1,400 years ago (1) due to their delicious high-quality meat. They are geographically distributed across the desert, wetland, and tropical forests (2). Rabbits are being reared at the domestic level as a livestock species for meat, wool, or fur production (3), and nowadays rabbit production has become a minor agricultural enterprise in Western

European countries such as Italy, Spain, and France (1, 3, 4). The worldwide per capita rabbit meat consumption is 0.242 kg, while in some European countries like Italy the per capita consumption is about 4.39 kg (5, 6). Furthermore, a stable rate of rabbit meat consumption has also been reported from 2000 to 2013 in many European countries (5).

The body growth is primarily regulated through complex interactive pathways that cause cell proliferation and cell enlargement (7). However, several factors including hormonal, genetic, environmental, and nutritional can remarkably predict the growth patterns (8). Myostatin (*MSTN*) is a member of the transforming growth factor (TGF-β) superfamily which is considered to be a negative regulator through inhibiting muscle development and regeneration (9). *MSTN* is an extracellular hormone that occurs in the skeletal muscle in an inactive state and restores its activity by binding to the precursor peptide, follistatin 3, and TGF-β binding protein, and transmitting the signals through the receptors (9, 10). The binding affinity of *MSTN* and ActRIIB could activate a chain of signal transmission that inhibits the myocyte differentiation and proliferation (10–12).

The *MSTN* gene inactivation might have an effect on muscle development and regeneration since *MSTN* gene deletion in mice resulted in a double muscle phenomenon (hyperplasia or hypertrophy) and a significant increase in muscle mass (10–12). Natural mutations of *MSTN,* which have an obvious double muscle phenotypic effect were found in cattle (13, 14), dogs (15), sheep (16, 17), pig (18), and humans (19, 20). CRISPR/Cas9 system is an efficient genome editing tool and has widely been used for functional gene study. Knock out of *MSTN* with CRISPR/Cas9 system has been processed in sheep (21, 22), goats (23, 24), and pigs (23, 25). Qian et al. have reported the *MSTN*-KO gene in Meishan pig fetal fibroblasts by engineered zinc-finger proteins and prepared the *MSTN*-KO pig by somatic cell nuclear transfer technology. In comparison to wild-type Meishan pigs, the lean meat rate of Meishan pigs with *MSTN* knockout increased by 11.62% (26). In addition, *MSTN*-KO goats' meat production was 32% higher than the wild-type (27).

Rabbits offer quick breeding sources and meat with excellent nutritive and dietetic properties which contain high-quality protein and low fat and cholesterol contents, that is fine-grained white meat which can substitute chicken, and is also ideal for obese and cardiovascular patients (28). In the present study, we have designed a highly efficient novel regulatory site and additionally verified that the exon 3 region can regulate the *MSTN* gene. Furthermore, in our study, the rabbits obtained after the *MSTN* gene editing are able to pass the gene-edited traits to the next-generation normally, and homozygous gene-edited rabbits obtained in the F2 generation have normal production performance and stable double muscle buttock characteristics.

Abbreviations: CRISPR, Clustered regularly interspaced short palindromic repeats; sgRNAs, small guide RNA; FBS, Calf Serum; DMEM, Dulbecco's modified eagle medium.

MATERIALS AND METHODS

Animals

In the present study, rabbits were used to perform all the experiments according to the Principle Guideline for the Use and Care of Laboratory Animals, Guangxi University. Rabbits were kept under controlled conditions at the Animal Center of Guangxi University. A total of 33 rabbits used for ovulation were approximately 6–8 months old and weighed between 3.5 and 4.5 kg.

SgRNA Design and Plasmid Construction

A total of 8 sgRNAs *MSTN* sites were designed using the website (http://www.genome-engineering.org/) and named as g1, g2, g3, g4, g5, g6, g7, and g8 (**Figure 1A**). The complementary DNA strand was annealed to develop into a double strand and cloned into the pMD18-hU6-gRNA vector. Primers T76 and T78 were used as marked gene fragments and amplified by targeting two sgRNA vectors. Linearized DNA plasmids and PCR products were extracted and purified using the MEGA shortscrip TM T7 kit (Ambion, USA). The sgRNAs were generated according to the manufacturer's recommendations.

Fibroblasts Culturing and Electroporation

Primary rabbit fibroblasts were cultured in 10% (v/v) FBS, 35 mm glutamine, and 1 × DMEM (Gibco). The plasmid was transfected into fibroblasts through electroporation (Gene Pulser Xcell & BIO-RAD) at a pulse of 225 volts for 10 ms. Briefly, 4 μg of pCMV-T7-NLS-hSpCas9-NLS (hSpCas9) DNA plasmid and 2 μg of gRNA plasmid were used to transfect 5×10^5-10×10^5 fibroblasts. After 24 h the culturing media was changed to DMEM containing 10% FBS. Subsequently cultured for 4 days, the DNA was extracted and the *MSTN* was amplified by PCR and sent to the Beijing Genomics Institute (BGI) for sequencing to analyze the editing effect of the gene fragment. Optimal sgRNA was selected for the following experiment.

Zygote Collection and Blastocyst Genotype Testing

Rabbits ($n = 3$) in estrus were mated, the female rabbits were anesthetized 16–20 h after mating, and the fertilized eggs were collected from the oviduct with a cell culture medium consisting of M199 (Gibco, USA), 3% bovine serum (Gibco, USA), 5 mM HEPES (Sigma, USA), and 5 mM NaHCO3 (Sigma, USA). Then, the zygotes were microinjected with Cas9 protein and sgRNA and the blastocysts were cultured *in vitro*. The mutation of the target region was analyzed by PCR and T7EI digestion assay.

Generation of Transgenic Rabbit

The zygote cytoplasm was injected with a 10-pl CRISPR/Cas9 mixture containing 200 ng/μl Cas9 protein (A36497, Hot Fisher, USA), 20 ng/μl sgRNA6, and 20 ng/μl sgRNA8, and the zygotes were transferred back to the oviduct of the recipient rabbit ($n = 30$). The pregnancy phenomenon was checked 10 days after the transplantation and the pups were born after 30 days.

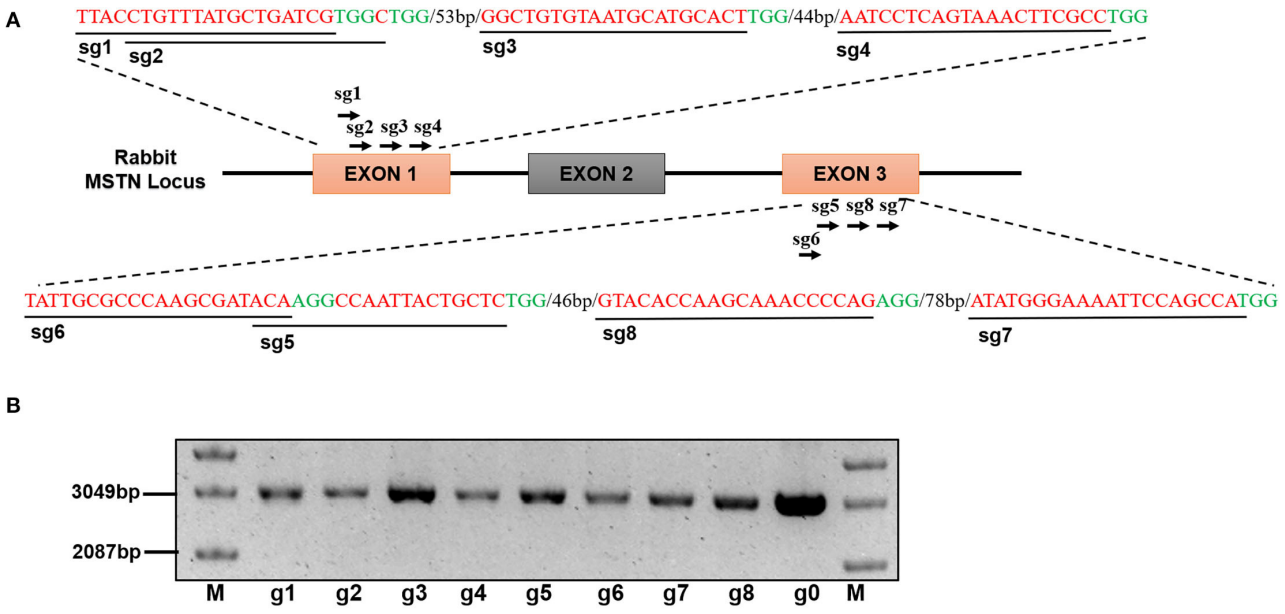


FIGURE 1 | (A) Design and construction of sgRNA Schematic diagram of sgRNA targeting site in the rabbit *MSTN* gene locus. **(B)** Electropherogram of eight sgRNA vectors, M: Supercoiled DNA Ladder Marker, g1-g8 refer to phU6-gRNA-Sg1to phU6-gRNA-Sg8 (The picture has been cropped).

T7 Endonuclease I (T7EI) Assay

Rabbit ear tissue genomic DNA was extracted using an extraction kit (Quick-DNA Plus, ZYMO RESEARCH) and CRISPR/Cas9-induced mutations were studied by PCR and T7E1 assay. PCR was performed using primers flanking the target site and the product was sequenced (Shanghai Sangon Company, Shanghai, China) and digested with T7EI (NEB, USA) to characterize the mutation.

Off-Target Analysis

Potential off-target sites of the sgRNAs were assessed using the CRISPR design tool (http://crispor.tefor.net/) and the top 4 potential off-target sites of each sgRNA were selected and amplified by PCR with specific primers (**Table 1**). PCR products were evaluated by sequencing and T7E1 digestion assay.

Western Blot Analysis

Gluteus maximus tissue samples were collected from 5 *MSTN*-KO and 5 WT rabbits (anesthesia sampling at 6 months of age), the tissues were ground in liquid nitrogen and 2.5 μl/ml protease inhibitor was added and kept in ice for 30 min. Protein concentration was determined using the Bradford method (Bio-Rad), and 35 μg of protein sample was subjected to a 5–12% separation by SDS-polyacrylamide gel. Anti-MSTN polyclonal antibody (Abcam) and goat anti-rabbit IgG were coupled to horseradish peroxidase (HRP; Santa Cruz, USA) antibodies for protein detection, while β-actin antibody (Santa Cruz) was used as an internal control.

TABLE 1 | Primers of off-target detection.

Primer	Sequence (5′ → 3′)	Length (bp)
POTs6-1F	CAATGGTGTGAGCCTCAAAG	382
POTs6-1R	AGTGGTCGTCTTCTTCATCC	
POTs6-2F	ATGCTCCTGTGTAGTCACTG	373
POTs6-2R	GTTTTCCATGTCCAGCTCAC	
POTs6-3F	ATCCAGGTATTAGCAACCGT	292
POTs6-3R	TCTGTGAATGTGCATACATACA	
POTs6-4F	TGACTACTGGCCCAAAATGT	450
POTs6-4R	TTCAGTCACAGAGTCGGTTT	
POTs6-5F	GAAGGGCCACAAAGAGAAAG	238
POTs6-5R	AAGGCCTCTTCCTCCCAG	
POTs8-1F	GGCGCTCATGATCTCTTGCT	291
POTs8-1R	CCTCACCAATGTCGATGCCT	
POTs8-2F	AGCAGACATTCTGGCGGAAA	348
POTs8-2R	GCCAAATGCAGCCTCAGAAT	
POTs8-3F	TCCAGAGTGCCTGCAGGTA	351
POTs8-3R	TCCAGAAGCTCAAATCTCTTGCT	
POTs8-4F	GTTTGGACACTGCTTGCTGG	291
POTs8-4R	TCATGGTGGATGCCCTCTTG	
POTs8-5F	CACACACATCCTCGGCTCAT	352
POTs8-5R	CATTACCTTCCTCCCACCCC	

Sample Collection, Physical and Histological Analysis

A total of 10 rabbits, including 5 F0 *MSTN*-KO and 5 WT, were fed at the same conditions and weaned on the 30th day. Rabbits were kept in separate cages and their body weight was recorded

after every 10 days for 200 days. At the age of 6 months, the rabbits were anesthetized with pentobarbital sodium (1 mL/kg) to take gluteal muscle mass tissue. The gluteus maximus tissues of the F0 generation *MSTN*-KO and WT rabbits were fixed with 4% paraformaldehyde at 4°C. Different concentrations of dehydrated sucrose (30% for 7 h, 40% for 7 h, 45% for 5 h) were also used and frozen at −25°C for histological examination. The tissue slices of 5 μm thickness were first stained with hematoxylin and eosin (H&E) and then examined by a fluorescence inverted microscope (Nikon, Japan). The integral optical density (IOD) analysis of histological sections was performed by using Image-Pro Plus 6.0 software.

Growth Data Recording of Edited Rabbit F1 and F2 Generations

The F0 generation edited rabbits were bred with wild type to obtain the F1 generation rabbits with the edited genotype. There were 3 male parents of edited rabbits and 22 wild-type female rabbits. All the F1 generation-produced edited rabbits were well developed and had significant differences in body weight from the wild type and were reserved for seed use. After sexual maturity, two breeding methods were adopted for breeding with the same genotype, and the edited male rabbits were bred with wild-type female rabbits to produce the F2 generation. Among them, there were 4 male-edited rabbits, 5 female-edited rabbits, and 9 wild-type female rabbits of the F1 generation. Genotype identification was carried out and two genotypes insertion type and deletion type were selected. The F2 generation edited rabbits were crossbred with the same genotype and also with the wild type. The weight recording method is consistent with the F0 generation edited rabbit recording method. Among them, the breeding and farrowing information of the F1 generation male rabbits were recorded.

Statistical Analysis

All the data obtained in this study were analyzed using Graph pad prism software (*T*-test) and the $p < 0.05$ was considered statistically significant (29).

RESULTS

The SgRNA Designing and Construction

The exon 1 and 3 of the *MSTN* gene were targeted to design eight sgRNAs (g1, g2, g3, g4, g5, g6, g7, and g8) which were cloned into the pMD18T vector as shown in **Figure 1A**. The recombinant vector was named pMD18-hU6-gRNA and the sequence analysis results, as presented in **Figure 1B**, confirmed that the sgRNA has been successfully inserted into the vector.

Mutational Effect of the SgRNAs in Rabbit Fibroblasts

The rabbit primary fibroblasts were obtained from newborns (**Figure 2A**) and the mutation efficiency of the designed sgRNAs was confirmed. In g1, g2, g3, and g4 no mutations were generated while the rate of mutation in g5, g6, g7, and g8 was observed to be about 20–45% (**Figure 2B**). Further, the sgRNAs targeted to exon 3 were significantly better than that of exon 1. The gene alteration proportion of sgRNA6 (45%, 9/20) and sgRNA8 (40%, 8/20) were higher than that of the other gRNAs. Thus, sgRNA6 and sgRNA8 were selected for further experiments.

CRISPR/Cas9-Mediated MSTN Site-Specific KNOCK-OUT in Rabbit Zygotes

The CRISPR/Cas9 system containing sgRNA6 and sgRNA8 was injected into the zygotes to verify the site-specific deletion of the *MSTN* gene. A total of 16 zygotes were used for injection purposes, of which 10 developed to the blastocyst stage through *in vitro* maturation (**Figure 3A**). The genome-editing efficiency in the blastocysts was examined by PCR and Sanger sequencing, and the results revealed that 70% (7/10) of the blastocysts had genetic mutations (**Figure 3B**).

Generation of MSTN Knock-Out Rabbits

A total of 99 cytoplasmic injected zygotes were transplanted to the oviduct of 10 female receptor rabbits, and 23 rabbit pups were born after full-term gestation (**Table 2**). Gene-editing was evaluated by T7E1 digestion and Sanger sequencing analysis which presented that 5 of them were *MSTN* knock-out rabbits (**Figure 4**) and interestingly, three $MSTN^{+/-}$ rabbits were all males, while two $MSTN^{-/-}$ rabbits were all females. The further T7E1 and sequencing analysis showed that 5 *MSTN* knock-out rabbits were chimeras.

Off-Target Mutation Analysis of *MSTN* Knock-Out Rabbits

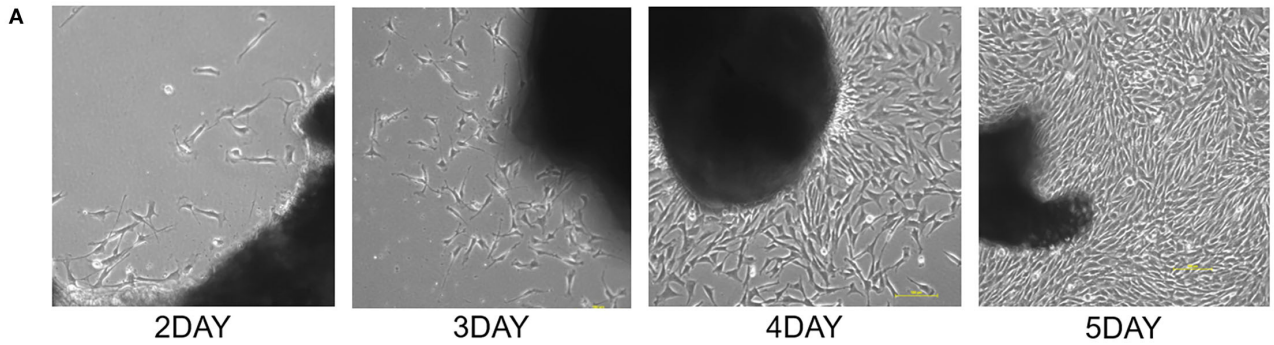
To detect the presence of off-target mutations in *MSTN* knockout rabbits, four potential off-target sites were designed. The T7EI digestion and Sanger sequencing results exhibited that no off-target mutation was detected in the *MSTN*-KO rabbits (**Figure 5**).

Western Blot Analysis of *MSTN*-KO Rabbits

According to the results of western blot analysis, it was revealed that the expression level of *MSTN* protein in the muscle of mutant rabbits was lower than that of WT rabbits (**Figure 6**).

Morphological Analysis of *MSTN*-KO Rabbits

The appearance of WT and F0 generation edited rabbits at the age of 3rd and 6th months were observed (**Figure 7A**). Under the same feeding conditions, the *biceps femoris* was more developed in the F0-generation edited rabbits than the WT. Moreover, the difference in muscle development of the *MSTN*-KO rabbit was evaluated using body weight, and data related to weight gain is shown in **Supplementary Table S1**. No significant difference was observed related to the body weight in the five *MSTN*-KO rabbits and the control group within the early 90 days ($p > 0.05$). But after growing to the age of 3rd month, the five *MSTN*-KO rabbits' weight began to appear slightly higher than that of the control group (**Figure 7B**) and the "double muscle" phenomenon appeared more obvious in *MSTN*-KO rabbits. At the age of 140 days, the five *MSTN*-KO rabbits exhibited a significantly higher body weight than the WT control group

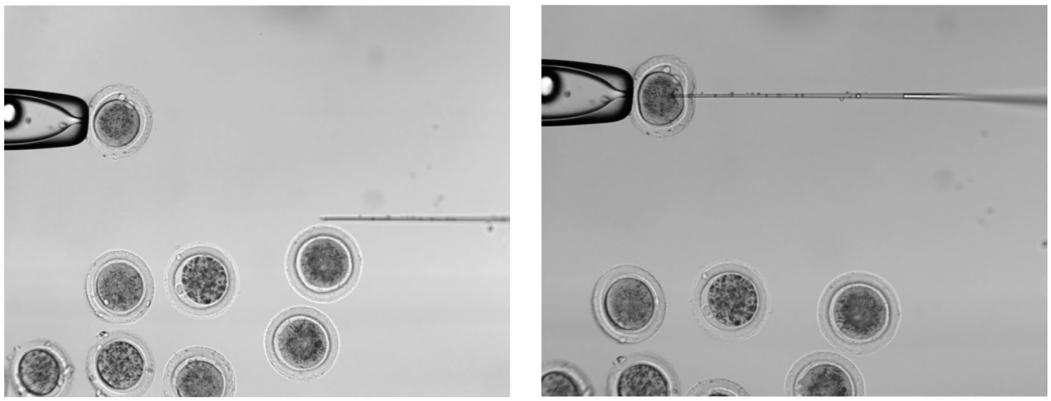


B

WT	AAGACGGGCTGTGTAATGCATGCACTTGGAGACAAAACAGTAAATAT	0
g1 （0，0%）	AAGACGGGCTGTGTAATGCATGCACTTGGAGACAAAACAGTAAATAT	0
g2 （0，0%）	AAGACGGGCTGTGTAATGCATGCACTTGGAGACAAAACAGTAAATAT	0
g3 （0，0%）	AAGACGGGCTGTGTAATGCATGCACTTGGAGACAAAACAGTAAATAT	0
g4 （0，0%）	AAGACGGGCTGTGTAATGCATGCACTTGGAGACAAAACAGTAAATAT	0
g5 （5/20，25%）	AAGAGATACAAGGCCAATTACTGCTCTGGAGAGTGTGAATTTGTG	0
	AAGAGATACAAGGCCAATTACT--CTCTGGAGAGTGTGAATTTGTG	-1
	AAGAGATACAAGGCCAATTACTGTTCTGGAGAGTGTGAATTTGTG	C-T
g6 （9/20，45%）	TTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACTGCTCTGG	0
	TTGGATTATTGCGCCCAAG-----TACAAGGCCAATTACTGCTCTGG	-3
	TTGGATTATTGCGCCCAAGAGATTACAAGGCCAATTACTGCTCTGG	+1
	TTGGATTATTGCGCCTAAGAGATACAAGGCCAATTACTGCTCTGG	C-T
	TTGGATTATTGCGCCCAAGAG-TACAAGGCCAATTACTGCTCTGG	-1
	TTGGATTATTGCGCCCAAGA-------CAAGGCCAATTACTGCTCTGG	-4
g7 （6/20，30%）	AATAATATATGGGAAAATTCCAGCCATGGTAGTAGACCGCTGTGG	0
	AATAATATATGGGAAAATTCCA--CCATGGTAGTAGACCGCTGTGG	-1
	AATAATATATGGGAAAATTCCA---- ---TGGTAGTAGACCGCTGTGG	-4
	AATAATATGTGGGAAAATTCCAGCCATGGTAGTAGACCGCTGTGG	A-G
	AATAATATATGGGAAAATTCCAGCCCATGGTAGTAGACCGCTGTGG	+1
	AATAATATACGGGAAAATTCCAGCCATGGTAGTAGACCGCTGTGG	T-C
g8 （8/20，40%）	CATCTTGTACACCAAGCAAACCCCAGAGGTTCAGCAGGTCCTTG	0
	CATCTTGTACACCAAGCAAACCC--AGAGGTTCAGCAGGTCCTTG	-1
	CATCTTGTACACCAAGCAAA----------AGGTTCAGCAGGTCCTTG	-6
	CATCTTGTACACCAAGCAAAC-----AGAGGTTCAGCAGGTCCTTG	-3
	CATCTTGTACACCAAGCA----------AGCAGGTCCTTG	-14

FIGURE 2 | The results of sequencing for mutations at the specific sites in rabbit REFs. **(A)** Rabbit Primary fibroblast cell in culture from 2days to 5days. **(B)** Sanger sequence analysis of REFs. g1-g4 target to exon1, g5-g8 target to exon 3. gRNA edited were named on the left. g1-g8 represent different gRNA, PAM sequence is marked in green. The sgRNAs sequence is marked in red. The numbers on the right indicated the type of mutation, with "-" represents deletion of the given number of nucleotides, "+" represents insertion of the given number of nucleotides.

rabbits ($P < 0.05$, $n = 5$). Furthermore, the bodyweight of *MSTN*-KO rabbits (3,572 ± 41.40 g) was significantly ($P < 0.01$) more at 160 days as compared to the control WT rabbit (3,094 ± 69.54 g). Additionally, the weight gain data from two selected genotypes (insertion and deletion type) of the F1 generation rabbit are shown in **Supplementary Table S2**. A significant weight gain has been observed for both the insertion type and the deletion type in comparison to WT at 140 days, where the bodyweight



B

sgRNA 6 sgRNA8

WT	GGGATTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	0
BL1,BL6,BL7	GGGATTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	0
BL2	GGGATTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	0
	GGGATTGGATTATTGCGCCCAAGAGA-ACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-1
	GGGATTGGATTATTGCGCCCAAGAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-7
	GGG----------CCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-30
BL3	GGGATTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	0
	GGGATTGGATTATTGCGCCCAAG---------- -------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-9
	GGGATTGGATTATTGCGCCCAA----------TAACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-15
	GGGATTGGATTATTGCGCCCAAGAGA--ACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-1
	GGGATTGGATTATTGCGCCCAATA----------ACT / / TTGTACACCAAGCAAACCC--AGAGGTTCAGC	-16
	GGGATTGGATTATTGCGCCCAA----------TAACT / / TTGTACACCAAGCA----------GAGGTTCAGC	-22
BL4,BL5	GGGATTGGATTATTGCGCCCAAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-9
	GGGATTGGATTATTGCGCCCAAG----------GCCAATTACT / / TTGTACACCAAGCAAAC--CCAGAGGTTCAGC	-10
	GGGATTGGATTATTGCGCCCAA----------TTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-15
BL8	GGGATTGGATTATTGCGCCCAAGAGATACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	0
	GGGATTGGATTATTGCGCCCAAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-9
	GGGATTGGATTATTGCGCCCAA----------TAACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-15
	GGGATTGGATTATTGCGCCCAAGAGA-ACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-1
	GGGATTGGATTATTGCGCCCAATA----------ACT / / TTGTACACCAAGCAAACCC--AGAGGTTCAGC	-16
	GGGATTGGATTATTGCGCCCAA----------TAACT / / TTGTACACCAAGCA----------GAGGTTCAGC	-22
BL9	GGGATTGGATTATTGCGCCCAAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-9
	GGGATTGGATTATTGCGCCCAAGAGA---ACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-1
	GGGATTGGATTATTGCGCCCAAGAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGC	-7
BL10	GGGATTGGATTATTGCGCCCAAGAGA-ACAAGGCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGCAGG	-1
	GGGATTGGATTATTGCGCCCAAG----------GCCAATTACT / / TTGTACACCAAGCAAACCCCAGAGGTTCAGCAGG	-9
	GGGATTGGATTATTGCGCCCAA----------ACT / / TTGTACACCAAGCA----------GAGGTTCAGCAGG	-22

FIGURE 3 | Targeted mutagenesis efficiency of the CRISPR/Cas9 system in rabbit zygotes. **(A)** Schematic diagram of the microinjection of Cas9 mRNA and sgRNAs into rabbit zygotes. **(B)** BL1-BL7 represent different blastocysts used in this study. Sanger sequence analysis of zygotes. zygotes edited were named on the left. PAM sequence is marked in green. The sgRNAs sequence is marked in red. The numbers on the right indicated the type of mutation, with "-" represents deletion of the given number of nucleotides.

of insertion type rabbits was 3,398 ± 92.97 g, deletion type rabbits 3,554 ± 110.45 g, and WT rabbits 3,064 ± 80.35 g ($P < 0.05$, $n = 5$), respectively. The *gluteus* of the *MSTN*-KO rabbits was also bigger than the WT rabbits, and histological analysis showed that in *MSTN*-KO rabbits the density and diameter of the myofiber were significantly large (**Figure 7C**). Also, it was found that the phenotypic differences of the edited rabbits were stably inherited in the F2 generation, and **Figure 7D** shows three young rabbits in the same litter. Similarly, T7E1 enzyme digestion and sequencing identification was used for young rabbit genotyping, and it has been observed that the body size of the edited rabbit was significantly different from that of the WT.

TABLE 2 | Generation of the *MSTN*-KO rabbits via CRISPR/Cas9.

No.	SgRNA	gRNA/Cas9 protein (ng/uL)	Embryos injected	Embryos transferred	Pregnancy	Pups obtained (%transferred)	*MSTN* KO pups (%pups)
M02	Sg6+Sg8	40/200	14	14	Yes	3 (21.40%)	1 (33.33%)
M08	Sg6+Sg8	40/200	12	12	Yes	3 (25.00%)	1 (33.33%)
M09	Sg6+Sg8	40/200	7	7	Yes	4 (57.14%)	0 (0)
M10	Sg6+Sg8	40/200	10	10	Yes	4 (40.00%)	2 (50.00%)
M12	Sg6+Sg8	40/200	9	8	Yes	4 (50.00%)	1 (25.00%)
M13	Sg6+Sg8	40/200	8	8	Yes	3 (37.50%)	0 (0)
M14	Sg6+Sg8	40/200	7	14	Yes	2 (14.29%)	1 (50.00%)
M15	Sg6+Sg8	40/200	9	7	Yes	4 (57.14%)	1 (25.00%)
M17	Sg6+Sg8	40/200	11	11	Yes	4 (36.36%)	2 (50.00%)

Heritability of the *MSTN*-KO Rabbits

To determine whether the offspring could stably inherit the successfully edited gene fragments from the *MSTN*-KO, all the F0 generation edited rabbits were crossed with wild-type rabbits. The three *MSTN*-KO F0 male rabbits were mated with wild rabbits and live F1 rabbit pups were produced. The T-cloning sequencing and T7E1 cleavage assay demonstrated that 28 out of the 53 newborn F1 rabbits carried *MSTN* mutations, of which 23 were monoallelic, and 5 were biallelic *MSTN*-KO rabbits (**Figure 8**). Two genotypes, deletion type and insertion type, were selected from the F1 generation to continue breeding, and the edited rabbits of the F2 generation were obtained. There was no significant difference in the litter size of female rabbits after breeding between F1 generation male rabbits and wild-type female rabbits, and there was no significant difference in the birth weight of offspring (**Figure 8**). So, there were three homozygous edited rabbits at the age of 1 month, with obvious differences in appearance and body shape. As shown in **Figure 8**, it is expected that a small number of homozygous edited rabbits obtained from the F2 generation, and in the F3 generation these edited homozygous rabbits continue to produce until a stable expansion thereof be achieved.

DISCUSSION

Humans are entirely reliant on livestock for their daily food supply, which comes in the form of eggs, meat, and milk (30–33). Thus, genetic alteration offers an opportunity for significant production and gains in a short time (34, 35). Since the last decade, many gene-edited organisms have been approved and are being used by the public, such as the gene-edited goats' milk and chicken eggs are being used for drug extraction. Further, the salmon was the first gene-edited species to be approved to be consumed as food, and recently the GalSafe pigs have also been certified to be used for food and medical purpose (36). Rabbit farming is now very well developed all around the world and systematically reared on a large scale, and the global rabbit meat production is reaching 1.8 million metric tons per year (37). As high-quality meat, the demand for rabbit meat is worth looking forward to in the future, but rabbits' muscular development is still a challenge. Therefore, *MSTN* gene-edited rabbits have the possibility of being approved and used by the public under the advancement of our multi-generation continuous observational research. In the present study, we used CRISPR-Cas9 to delete a long gene fragment of the *MSTN* gene, which ensured the inactivation of target gene function, and the results of our study presented a successful production of *MSTN*-KO rabbit with heritable ability, and the loss of the *MSTN* fragment could lead to muscle growth. In both the F0 and F1 generations, there were significant differences in body weight before it reached the plateau, and the average body weight of the edited rabbits before the slaughter age was also higher than that of the WT rabbits.

Although the increase in muscle mass makes *MSTN*-KO livestock production more attractive, in these *MSTN* mutant animals calving difficulties are often caused by various reasons, such as large offspring syndrome (LOS) has been reported in *MSTN*-deficient animals (15, 38). Compared with other livestock, *MSTN* gene-edited rabbits rarely report dystocia due to large fetuses, even though in our study the *MSTN*-KO rabbits exhibited a typical double-muscle phenotype with increased body weight but, at birth, no significant differences have been observed in body weight and size as compared with the control WT. Further, the *MSTN*-KO rabbits appeared healthy and normal without reproduction difficulties, demonstrating that CRISPR/Cas9 system-generated *MSTN*-KO rabbits are best suited for studying muscle development and associated diseases, and the results of our study are in line with those of Lv et al. (39).

MSTN gene can code for TGF-beta (transforming growth factor-beta) superfamily ligand that can bind with different TGF-beta receptors and ultimately recruit or activate the SMAD family transcription factors, which regulate the gene expression (40). *MSTN* is a key regulator of muscle cell proliferation and differentiation throughout muscle development, and *MSTN* expression signals were detected from the early myogenic stage of embryonic sarcomere formation to the adult skeletal muscle development stage (14, 41, 42). Earlier, the birth weight of *MSTN* mutant animals has also been reported in sheep, cattle, and goats (43, 44). Wang et al. have also described the *MSTN* gene-modification and their effects in goats with larger muscle fiber size resulting in enhanced bodyweight (43). Therefore, the *MSTN* gene knockout might affect muscular development

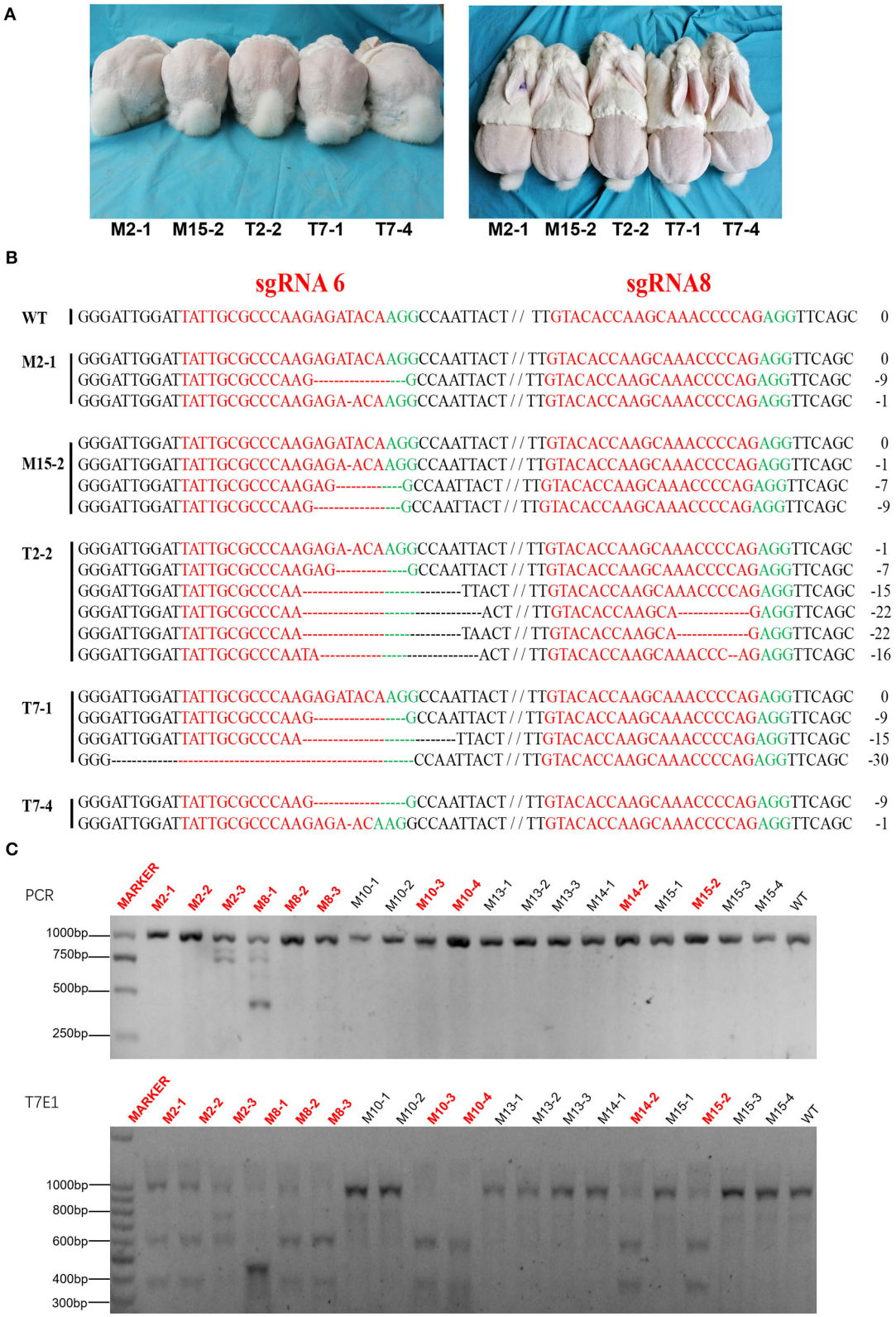


FIGURE 4 | Generation of F0 *MSTN* KO rabbits via CRISPR/Cas9. **(A)** Five mutated rabbits of F0. **(B)** Sanger sequence analysis of mutated rabbits. Five rabbits edited were named on the left. WT stands for wild type control; PAM sequence is marked in green. The sgRNAs sequence is marked in red. The numbers on the right indicated the type of mutation, with "-" represents deletion of the given number of nucleotides. **(C)** Upper: The results of sequencing for mutations at the specific sites in Agarose gel electrophoresis of PCR products about a part of pups of F0. Lower: T7E1 cleavage assay for the rabbits (The picture has been cropped).

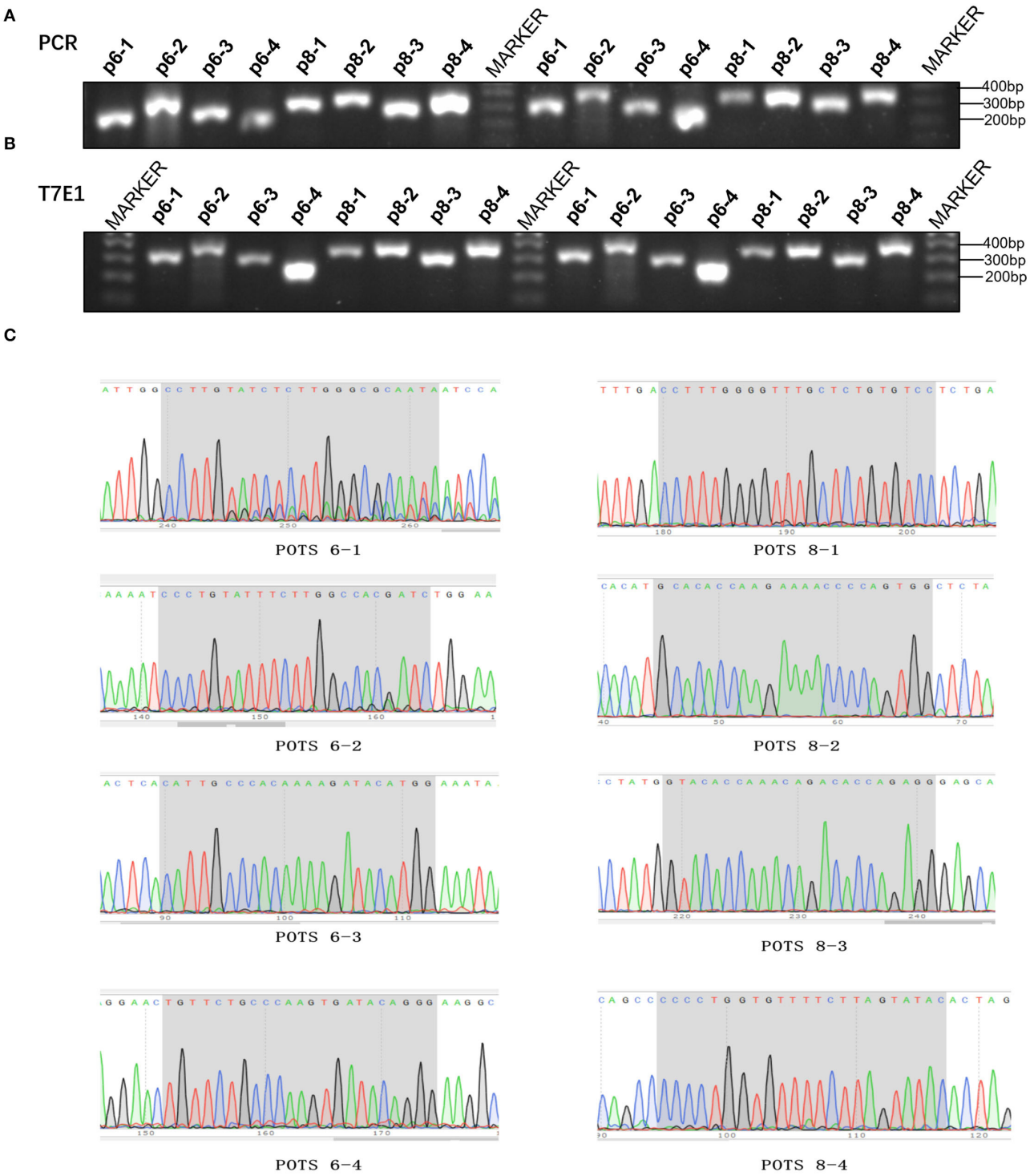


FIGURE 5 | Detection of *MSTN* modified rabbits. **(A)** The results of suspicious POTs in Agarose gel electrophoresis of PCR products. M: DL100bp; p6-1–p6-4 represent g6 four suspicious POTs, p8-1–p8-4 represent g8 four suspicious POTs (The picture has been cropped). **(B)** The results of suspicious POTs in Agarose gel electrophoresis of T7E1. **(C)** Sanger sequencing for suspicious POTs sites.

at the embryonic stage (43, 44). Furthermore, it has been stated that *MSTN* can only express in skeletal muscle during fetal development, thereby controlling the differentiation and proliferation of myoblasts (43, 44). In our study, *MSTN*-KO

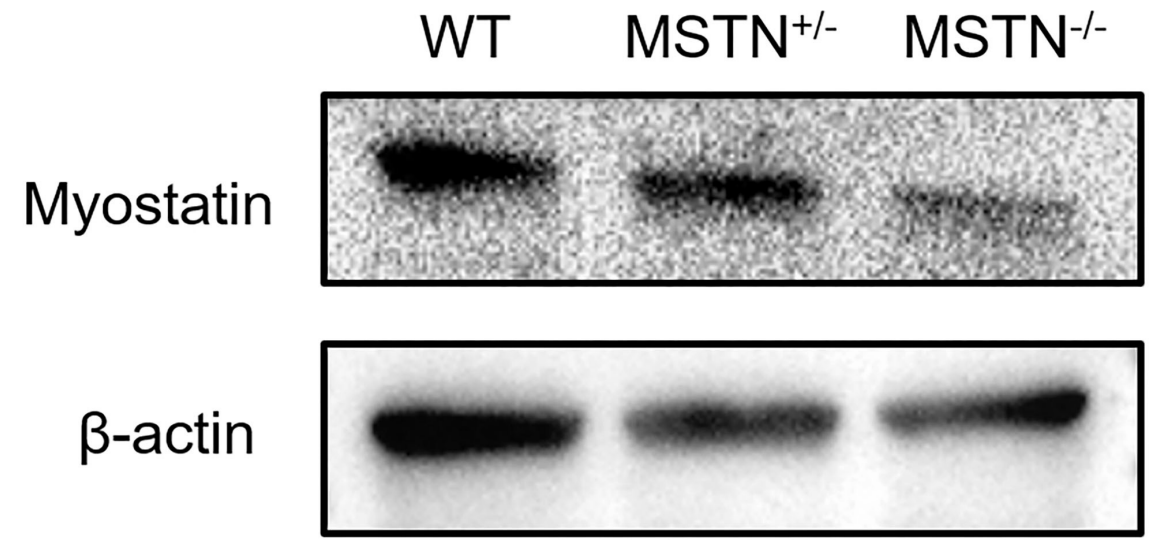


FIGURE 6 | Result of *MSTN+/–, MSTN–/–* and WT rabbits in Western Blot. Beta-actin served as a loading control (The picture has been cropped).

rabbits exhibited a typical double-muscle phenotypic trait, but there was no significant difference in body size at birth ($P > 0.05$) and body weight in the first 60 days compared with the WT group. This might be because of the IGF-1 which is an important positive regulator of muscle cell proliferation and differentiation in the skeletal muscle (45). The IGF1 signaling pathway is critically involved in long-term health regulation, which ultimately plays an essential role to control various homeostatic mechanisms related to growth or development (46). Despite the inhibition of *MSTN* signal expression, IGF signaling still might upregulate the expression of myostatin in skeletal muscle tissue models, which indicated the presence of an autoregulatory inhibitory loop in a muscular system (47, 48).

Moreover, the muscle fiber hypertrophy, hyperplasia, or a combination thereof triggered by the lack of the *MSTN* gene subsequently results in increased mammalian muscle production (49–52). The quantity of muscle fibers is mainly determined before birth because the diameter of muscle fibers might be increased after birth (53, 54). The mechanism of muscle hypertrophy and hyperplasia may not occur simultaneously, and it appears that the mechanism by which *MSTN* mutations enhance the muscle mass varies among species (55, 56). In *MSTN* mutant mice and goats, the muscle mass gain was caused by muscle fiber hyperplasia and diameter hypertrophy (39, 52, 57), while only muscle fiber hyperplasia was observed in *MSTN* mutant cattle and pigs (13, 50). The results of the current study presented that the *gluteus maximus* of the *MSTN*-KO rabbits was also bigger than the WT rabbits, and histological analysis showed that in *MSTN*-KO rabbits the density and diameter of the myofiber were significantly large. Therefore, we hypothesized that the enhanced muscular phenotype in the *MSTN*-KO rabbits is attributed to muscles fiber hyperplasia and/or hypertrophy. Furthermore, previously it has been reported that the muscle mass of *MSTN*-KO mice was 2–3 times more than that of the WT mice, which was the result of the combined action of muscle fiber hyperplasia and diameter hypertrophy (13, 52). On the other hand, the cattle with double-muscle trait showed that muscle mass increased by 20–25%, and the increased muscle weight appears to be the result of muscle fiber diameter proliferation, rather than an increase in muscle number (13, 50). In this study, compared with WT rabbits, the average body weight of F0 and F1 generation edited rabbits increased by 15 and 15.99% at the age of 180 and 140 days, respectively. The fragment deletion targeting a specific site in *MSTN*-KO rabbits could suggest that the sgRNA-based CRISPR/Cas9 system might be a useful tool for gene knockout in the mammalian genome (58). Earlier in rabbits, Lv et al. (39) designed two knockout sites at the first exon to edit the *MSTN* gene (39), while we designed a total of 8 sites in two exons (exon 1 and exon 3) each with four sites. After selection, we determined that the knockout efficiency of the third exon is better than the first exon, thus, we obtained edited rabbits with different knockout sites from already reported *MSTN* knockout studies. Moreover, at present, most of the experimental research regarding animals' gene-editing is still in the F0 generation (24, 59, 60), including the reported gene editing in rabbits (39). While, no study has yet been performed on the F1 or their subsequent generations, we not only developed *MSTN* gene-edited rabbits in the F0 generation but also continued the follow-up research in successive F1 and F2 generations with stable heritable mutant populations. Also, the difference between the body type of mutant and the wild type can be observed within 1 month after birth. Gene-editing technology can accelerate the breeding process and improve the meat production of livestock, but the congenital defects of gene edited animals and the uncertainty of trait stability have a great

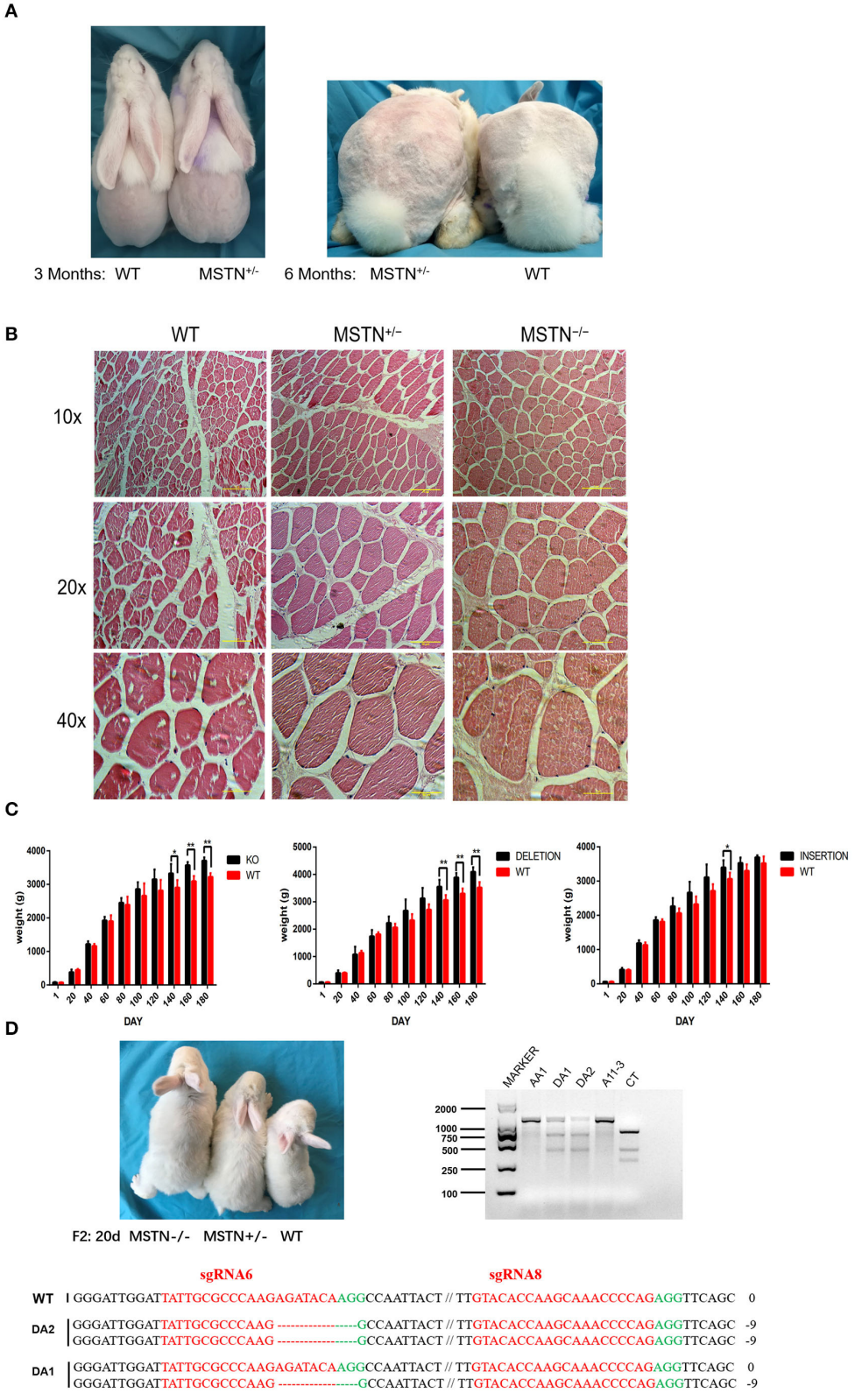


FIGURE 7 | Detection and observation of *MSTN*-KO rabbits. **(A)** Comparison of mediated and wild-type rabbits at 3 and 6 months of age. **(B)** The average body weight of $MSTN^{+/-}$ and WT rabbits from F0 and F1 ($n = 5$), and the left side shows the weight gain of F0 knockout rabbits and wild rabbits, the middle shows the weight gain of F1 deletion editing rabbits and wild rabbits, and the right side shows the weight gain of F1 insertion editing rabbits and wild rabbits. **(C)** HandE staining of the muscle fibers from *gluteus maximus*. **(D)** Shape difference map and identification results of F2 generation rabbits at the age of 20 days (The picture has been cropped).

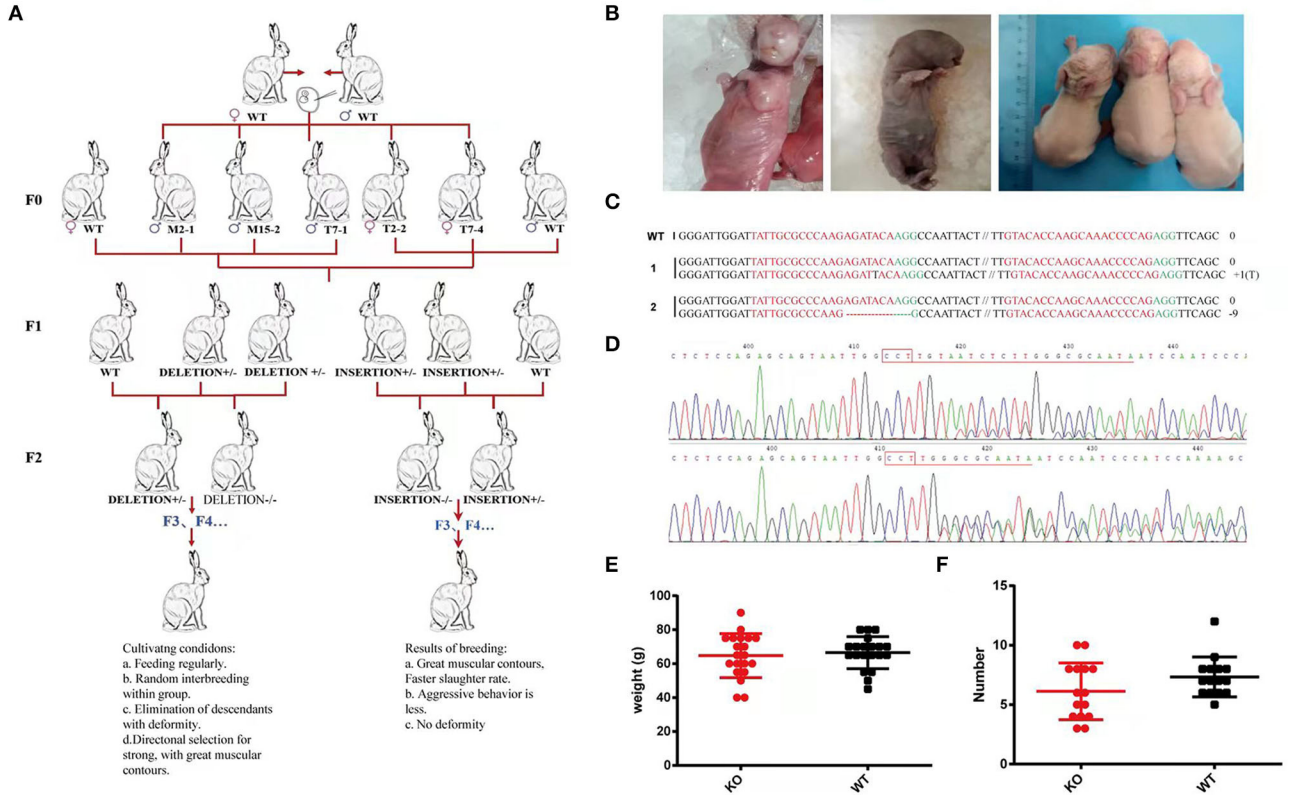


FIGURE 8 | (A) Schematic diagram of *MSTN*-KO rabbits breeding family. **(B)** Left: A cub with a large tongue. Right: Difficult to give birth due to the large size of the cubs. **(C)** Rabbits of F1 were named on the left. 1-2 represent different rabbits, PAM sequence is marked in green. The sgRNAs sequence is marked in red. The numbers on the right indicated the type of mutation, with "-" represents deletion of the given number of nucleotides, "+" represents insertion of the given number of nucleotides. **(D)** sequence identification. **(E)** Birth weight of rabbits. **(F)** Litter size.

impact on the commercialization of edited animals (51, 61). In our study medium- and long-term breeding can greatly reduce concerns about the phenotypic and genetic stability of *MSTN* gene-edited animals.

Moreover, various studies reported different types of skeletal-related problems in gene-edited animals such as *MSTN* gene-edited rabbits had bone deformities, pelvic tilt, and tooth dislocation, while *MSTN* gene knock-out Meishan pigs had an extrathoracic vertebra (26, 37). Abnormal pelvic structure may cause difficulties in calving in female animals. But, in our study, the *MSTN* gene-edited rabbits obtained by knocking out the third exon did not show these symptoms in appearance, which are more suitable for the breeding of strains. Additionally, in the previous study, the *MSTN*-KO animals exhibited the muscular feature, but the *MSTN*-KO caused severe health issues, such as the fetus having "large tongue syndrome," which results in milk suckling problems and death (52). In our study, the *MSTN*-KO rabbits not only exhibited double mushy buttocks but were also able to inherit this trait to subsequent generations without any problem. The majority of *MSTN*-KO rabbits showed a rapid increase in muscle mass after the age of 60 days, and the *MSTN*-KO rabbits were healthy and could inherit the knocked-out gene fragments, demonstrating that the *MSTN*-KO rabbits produced by the CRISPR/Cas9 system are suitable for studying muscle development and related diseases.

CONCLUSION

The rabbit meat is an enriched source of minerals with low fat and calories and is easy to digest, making it valuable for the consumers, especially for persons suffering from cardiovascular diseases or obesity. Taken together, we developed *MSTN*-KO rabbits with a typical phenotypic trait of double muscle buttocks that are likely to obtain edited rabbits lines which would help to improve the rabbit's meat production efficiency and promote the development of the rabbit industry. Furthermore, these *MSTN*-KO rabbits could be a promising tool for studying the development of muscles in other livestock species and improving their important economic trait.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in Genbank repository accession number MZ427879-MZ427887 at: https://www.ncbi.nlm.nih.gov/nuccore/.

ETHICS STATEMENT

The animal study was reviewed and approved by Experimental Animal Care and Use Committee of Guangxi University (Permit Code: 2022-GXU-006). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

SR, KC, and QL: conceptualization and resources. YZ, YLZ, and LW: data curation. YLZ, YZ, and ZL: methodology. HR, ZL, and DS: software. QL and KC: supervision. YZ and YLZ: writing—original draft preparation. YLZ, YZ, HR, ZL, DS, SR, QL, and KC: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

The present study was granted and supported by the National Natural Science Fund (Grant Nos. 31760648 and 31860638) and Guangxi Natural Science Foundation (Grant No. AB16380042).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.842074/full#supplementary-material

REFERENCES

1. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science.* (2014) 345:1074–9. doi: 10.1126/science.1253714
2. Smith, A. T. (2021). "rabbit". Encyclopedia Britannica. (2021) Available online at: https://www.britannica.com/animal/rabbit. (accessed 24 October 2021).
3. Dalle Zotte A. Rabbit farming for meat purposes. *Animal Front.* (2014) 4:62–7. doi: 10.2527/af.2014-0035
4. Patton N, Cheeke P. *The rabbit as a meat producing animal* [Feeding, meat composition]. In: *Paper presented at the Proceedings Annual Reciprocal Meat Conference.* (1981).
5. Gomer RH. Not being the wrong size. *Nat Rev Molec Cell Biol.* (2001) 2:48–55. doi: 10.1038/35048058
6. Szendro K, Szabó-Szentgróti E, Szigeti O. Consumers' attitude to consumption of rabbit meat in eight countries depending on the production method and its purchase form. *Foods.* (2020) 9:654. doi: 10.3390/foods9050654
7. Lui JC, Garrison P, Baron J. Regulation of body growth. *Curr Opin Pediatr.* (2015) 27:502. doi: 10.1097/MOP.0000000000000235
8. Rodino-Klapac LR, Haidet AM, Kota J, Handy C, Kaspar BK, Mendell JR. Inhibition of myostatin with emphasis on follistatin as a therapy for muscle disease. *Muscle Nerve.* (2009) 39:283–96. doi: 10.1002/mus.21244
9. McCoy JC, Walker RG, Murray NH, Thompson TB. Crystal structure of the WFIKKN2 follistatin domain reveals insight into how it inhibits growth differentiation factor 8 (GDF8) and GDF11. *J Biol Chem.* (2019) 294:6333–43. doi: 10.1074/jbc.RA118.005831
10. Wicik Z, Sadkowski T, Jank M, Motyl T. The transcriptomic signature of myostatin inhibitory influence on the differentiation of mouse C2C12 myoblasts. *Polish J Veter Sci.* (2011) 14. doi: 10.2478/v10181-011-0095-7
11. Nomura T, Ueyama T, Ashihara E, Tateishi K, Asada S, Nakajima N, et al. Skeletal muscle-derived progenitors capable of differentiating into cardiomyocytes proliferate through myostatin-independent TGF-β family signaling. *Biochem Biophys Res Commun.* (2008) 365:863–9. doi: 10.1016/j.bbrc.2007.11.087
12. Sakamoto K, Kanematsu-Yamaki Y, Kamada Y, Oka M, Ohnishi T, Miwa M, et al. Identification of ligand-selective peptidic ActRIIB-antagonists using phage display technology. *Biochem Biophys Rep.* (2017) 11:33–9. doi: 10.1016/j.bbrep.2017.06.001
13. Grobet L, Martin LJR, Poncelet D, Pirottin D, Brouwers B, Riquet J, et al. A deletion in the bovine myostatin gene causes the double–muscled phenotype in cattle. *Nat Genet.* (1997) 17:71–4. doi: 10.1038/ng0997-71
14. Kambadur R, Sharma M, Smith TP, Bass JJ. Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Res.* (1997) 7:910–5. doi: 10.1101/gr.7.9.910
15. Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, et al. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* (2007) 3:e79. doi: 10.1371/journal.pgen.0030079
16. Boman IA, Våge DI. An insertion in the coding region of the myostatin (MSTN) gene affects carcass conformation and fatness in the Norwegian Spælsau (*Ovis aries*). *BMC Res Notes.* (2009) 2:1–5. doi: 10.1186/1756-0500-2-98
17. Sahu AR, Jeichitra V, Rajendran R, Raja A. Novel report on mutation in exon 3 of myostatin (MSTN) gene in Nilagiri sheep: an endangered breed of South India. *Trop Anim Health Prod.* (2019) 51:1817–22. doi: 10.1007/s11250-019-01873-7
18. Stinckens A, Luyten T, Bijttebier J, Van den Maagdenberg K, Dieltiens D, Janssens S, et al. Characterization of the complete porcine MSTN gene and expression levels in pig breeds differing in muscularity. *Anim Genet.* (2008) 39:586–96. doi: 10.1111/j.1365-2052.2008.01774.x
19. Schuelke M, Wagner KR, Stolz LE, Hübner C, Riebel T, Kömen W, et al. Myostatin mutation associated with gross muscle hypertrophy in a child. *New Engl J Med.* (2004) 350:2682–8. doi: 10.1056/NEJMoa040933
20. Saunders MA, Good JM, Lawrence EC, Ferrell RE, Li W-H, Nachman MW. Human adaptive evolution at Myostatin (GDF8), a regulator of muscle growth. *Am J Human Genet.* (2006) 79:1089–97. doi: 10.1086/509707
21. Li H, Wang G, Hao Z, Zhang G, Qing Y, Liu S, et al. Generation of biallelic knock-out sheep via gene-editing and somatic cell nuclear transfer. *Sci Rep.* (2016) 6:1–12. doi: 10.1038/srep33675
22. Zhang Y, Wang Y, Yulin B, Tang B, Wang M, Zhang C, et al. CRISPR/Cas9-mediated sheep MSTN gene knockout and promote sSMSCs differentiation. *J Cell Biochem.* (2019) 120:1794–806. doi: 10.1002/jcb.27474
23. Wang X, Yu H, Lei A, Zhou J, Zeng W, Zhu H, et al. Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9 system. *Sci Rep.* (2015) 5:1–9. doi: 10.1038/srep13878
24. Zhang J, Liu J, Yang W, Cui M, Dai B, Dong Y, et al. Comparison of gene editing efficiencies of CRISPR/Cas9 and TALEN for generation of MSTN knock-out cashmere goats. *Theriogenology.* (2019) 132:1–11. doi: 10.1016/j.theriogenology.2019.03.029
25. Xie S, Li X, Qian L, Cai C, Xiao G, Jiang S, et al. An integrated analysis of mRNA and miRNA in skeletal muscle from myostatin-edited Meishan pigs. *Genome.* (2019) 62:305–15. doi: 10.1139/gen-2018-0110
26. Qian L, Tang M, Yang J, Wang Q, Cai C, Jiang S, et al. Targeted mutations in myostatin by zinc-finger nucleases result in double-muscled phenotype in Meishan pigs. *Sci Rep.* (2015) 5:1–13. doi: 10.1038/srep14435
27. Zhang J, Cui ML, Nie YW, Dai B, Li FR, Liu DJ, et al. CRISPR/Cas9-mediated specific integration of fat-1 at the goat MSTN locus. *FEBS J.* (2018) 285:2828–39. doi: 10.1111/febs.14520
28. Dalle Zotte A, Szendro Z. The role of rabbit meat as functional food. *Meat Sci.* (2011) 88:319–331. doi: 10.1016/j.meatsci.2011.02.017
29. Duan Q, Li Y, Qin F. Research on the protection and mechanism of coptis alkaloids on acetic acid type gastric ulcer. *J Phys.* (2020) 1629:12028. doi: 10.1088/1742-6596/1629/1/012028
30. Rehman S, Shafique L, Yousuf MR, Liu Q, Ahmed JZ, Riaz H. Spectrophotometric calibration and comparison of different semen

evaluation methods in Nili-Ravi buffalo bulls. *Pak Vet J.* (2019) 39:568–72. doi: 10.29261/pakvetj/2019.073
31. Rehman SU, Nadeem A, Javed M, Hassan FU, Luo X, Khalid RB, et al. Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo. *Genes.* (2020) 11:1388. doi: 10.3390/genes11111388
32. Rehman SU, Feng T, Wu S, Luo X, Lei A, Luobu B, et al. Comparative genomics, evolutionary and gene regulatory regions analysis of casein gene family in bubalus bubalis. *Front Genet.* (2021) 12:662609. doi: 10.3389/fgene.2021.662609
33. Rehman SU, Hassan F-U, Luo X, Li Z, Liu Q. Whole-genome sequencing and characterization of buffalo genetic resources: recent advances and future challenges. *Animals Open Access J From MDPI.* (2021) 11:904. doi: 10.3390/ani11030904
34. Singh P, Ali SA. Impact of CRISPR-Cas9-based genome engineering in farm animals. *Veter Sci.* (2021) 8:122. doi: 10.3390/vetsci8070122
35. Wu S, Hassan F-U, Luo Y, Fatima I, Ahmed I, Ihsan A, et al. Comparative genomic characterization of buffalo fibronectin type III domain proteins: exploring the novel role of FNDC5/Irisin as a ligand of gonadal receptors. *Biology.* (2021) 10:1207. doi: 10.3390/biology10111207
36. Katiyar Kritika S, Burrell Justin C, Laimo Franco A, Browne Kevin D, Bianchi John R, Walters Anneke, et al. Biomanufacturing of axon-based tissue engineered nerve grafts using porcine galsafe neurons. *Tissue Eng Part A.* (2021) 27:1305–20. doi: 10.1089/ten.TEA.2020.0303
37. Zhang T, Lu Y, Song S, Lu R, Zhou M, He Z, et al. 'Double-muscling'and pelvic tilt phenomena in rabbits with the cystine-knot motif deficiency of myostatin on exon 3. *Biosci Rep.* (2019) 39:BSR20190207. doi: 10.1042/BSR20190207
38. Lee S-J. Regulation of muscle mass by myostatin. *Annu Rev Cell Dev Biol.* (2004) 20:61–86. doi: 10.1146/annurev.cellbio.20.012103.135836
39. Lv Q, Yuan L, Deng J, Chen M, Wang Y, Zeng J, et al. Efficient generation of myostatin gene mutated rabbit by CRISPR/Cas9. *Sci Rep.* (2016) 6:1–8. doi: 10.1038/srep25029
40. Alexopoulos T, Vasilieva L, Kontogianni MD, Tenta R, Georgiou A, Stroumpouli E, et al. Myostatin in combination with creatine phosphokinase or albumin may differentiate patients with cirrhosis and sarcopenia. *Am J Physiol Gastrointest Liver Physiol.* (2021) 321:G543–G551. doi: 10.1152/ajpgi.00184.2021
41. Manceau M, Gros J, Savage K, Thomé V, McPherron A, Paterson B, et al. Myostatin promotes the terminal differentiation of embryonic muscle progenitors. *Genes Devel.* (2008) 22:668–681. doi: 10.1101/gad.454408
42. Wang X, Niu Y, Zhou J, Zhu H, Ma B, Yu H, et al. CRISPR/Cas9-mediated MSTN disruption and heritable mutagenesis in goats causes increased body mass. *Anim Genet.* (2018) 49:43–51. doi: 10.1111/age.12626
43. Oldham J, Martyn J, Sharma M, Jeanplong F, Kambadur R, Bass J. Molecular expression of myostatin and MyoD is greater in double-muscled than normal-muscled cattle fetuses. *Am J Physiol-Regul Integr Comparat Physiol.* (2001) 280:R1488–93. doi: 10.1152/ajpregu.2001.280.5.R1488
44. Valdés JA, Flores S, Fuentes EN, Osorio-Fuentealba C, Jaimovich E, Molina A. IGF-1 induces IP3-dependent calcium signal involved in the regulation of myostatin gene expression mediated by NFAT during myoblast differentiation. *J Cell Physiol.* (2013) 228:1452–63. doi: 10.1002/jcp.24298
45. Yang W, Zhang Y, Li Y, Wu Z, Zhu D. Myostatin induces cyclin D1 degradation to cause cell cycle arrest through a phosphatidylinositol 3-kinase/AKT/GSK-3β pathway and is antagonized by insulin-like growth factor 1. *J Biol Chem.* (2007) 282:3799–808. doi: 10.1074/jbc.M610185200
46. Ziv E, Hu D. Ageing Res. *Rev.* (2011) 10:201–4. doi: 10.1016/j.arr.2010.09.002
47. Kurokawa M, Sato F, Aramaki S, Soh T, Yamauchi N, Hattori M-A. Monitor of the myostatin autocrine action during differentiation of embryonic chicken myoblasts into myotubes: effect of IGF-I. *Mol Cell Biochem.* (2009) 331:193–9. doi: 10.1007/s11010-009-0158-6
48. McPherron AC, Lawler AM, Lee S-J. Regulation of skeletal muscle mass in mice by a new TGF-p superfamily member. *Nature.* (1997) 387:83–90. doi: 10.1038/387083a0
49. McPherron AC, Lee S-J. Double muscling in cattle due to mutations in the myostatin gene. *Proc Nat Acad Sci.* (1997) 94:12457–61. doi: 10.1073/pnas.94.23.12457
50. Wegner J, Albrecht E, Fiedler I, Teuscher F, Papstein H-J, Ender K. Growth- and breed-related changes of muscle fiber characteristics in cattle. *J Anim Sci.* (2000) 78:1485–96. doi: 10.2527/2000.7861485x
51. Wang K, Ouyang H, Xie Z, Yao C, Guo N, Li M, et al. Efficient generation of myostatin mutations in pigs using the CRISPR/Cas9 system. *Sci Rep.* (2015) 5:1–11. doi: 10.1038/srep16623
52. Guo R, Wan Y, Xu D, Cui L, Deng M, Zhang G, et al. Generation and evaluation of Myostatin knock-out rabbits and goats using CRISPR/Cas9 system. *Sci Rep.* (2016) 6:1–10. doi: 10.1038/srep29855
53. Swatland H. Muscle growth in the fetal and neonatal pig. *J Anim Sci.* (1973) 37:536–45. doi: 10.2527/jas1973.372536x
54. Nishi M, Yasue A, Nishimatu S, Nohno T, Yamaoka T, Itakura M, et al. A missense mutant myostatin causes hyperplasia without hypertrophy in the mouse muscle. *Biochem Biophys Res Commun.* (2002) 293:247–51. doi: 10.1016/S0006-291X(02)00209-7
55. Zhu X, Hadhazy M, Wehling M, Tidball JG, McNally EM. Dominant negative myostatin produces hypertrophy without hyperplasia in muscle. *FEBS Lett.* (2000) 474:71–5. doi: 10.1016/S0014-5793(00)01570-2
56. He Z, Zhang T, Jiang L, Zhou M, Wu D, Mei J, et al. Use of CRISPR/Cas9 technology efficiently targetted goat myostatin through zygotes microinjection resulting in double-muscled phenotype in goats. *Biosci. Rep.* (2018) 38. doi: 10.1042/BSR20180742
57. Bi Y, Hua Z, Liu X, Hua W, Ren H, Xiao H, et al. Isozygous and selectable marker-free MSTN knockout cloned pigs generated by the combined use of CRISPR/Cas9 and Cre/LoxP. *Sci Rep.* (2016) 6:1–12. doi: 10.1038/srep31729
58. Zhou J, Wang J, Shen B, Chen L, Su Y, Yang J, et al. Dual sgRNAs facilitate CRISPR/Cas9 mediated mouse genome targeting. *FEBS J.* (2014) 281:1717–1725. doi: 10.1111/febs.12735
59. Dingwei P, Ruiqiang L, Wu Z, Min W, Xuan S, Jianhua Z, et al. Editing the cystine knot motif of MSTN enhances muscle development of Liang Guang Small Spotted pigs. *Yi Chuan.* (2021) 43:261–70. doi: 10.16288/j.yczz.20-222
60. Gim GM, Kwon DH, Eom KH, Moon J, Park JH, Lee WW, et al. Production of MSTN-mutated cattle without exogenous gene integration using CRISPR-Cas9. *Biotechnol J.* (2021) 11:e2100198. doi: 10.1002/biot.202100198
61. Kang Q, Hu Y, Zou Y, Hu W, Li L, Chang F, et al. Improving pig genetic resistance and muscle production through molecular biology. In: *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*. Vancouver. (2014).

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

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

frontiers | Frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/fvets.2022.892815

Genome-Wide Association Studies, Runs of Homozygosity Analysis, and Copy Number Variation Detection to Identify Reproduction-Related Genes in Bama Xiang Pigs

Jiayuan Mo[1], *Yujie Lu*[1], *Siran Zhu*[1], *Lingli Feng*[1], *Wenjing Qi*[1], *Xingfa Chen*[1], *Bingkun Xie*[1,2], *Baojian Chen*[2], *Ganqiu Lan*[1] *and Jing Liang*[1]*

[1] College of Animal Science & Technology, Guangxi University, Nanning, China, [2] Guangxi Key Laboratory of Livestock Genetic Improvement, Guangxi Institute of Animal Science, Nanning, China

OPEN ACCESS

Edited by:
Yang Zhou,
Huazhong Agricultural University, China

Reviewed by:
Kejun Wang,
Henan Agricultural University, China
Majid Khansefid,
La Trobe University, Australia

***Correspondence:**
Jing Liang
liangjing@gxu.edu.cn

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

Received: 09 March 2022
Accepted: 22 April 2022
Published: 31 May 2022

Citation:
Mo J, Lu Y, Zhu S, Feng L, Qi W, Chen X, Xie B, Chen B, Lan G and Liang J (2022) Genome-Wide Association Studies, Runs of Homozygosity Analysis, and Copy Number Variation Detection to Identify Reproduction-Related Genes in Bama Xiang Pigs. Front. Vet. Sci. 9:892815. doi: 10.3389/fvets.2022.892815

Litter size and teat number are economically important traits in the porcine industry. However, the genetic mechanisms influencing these traits remain unknown. In this study, we analyzed the genetic basis of litter size and teat number in Bama Xiang pigs and evaluated the genomic inbreeding coefficients of this breed. We conducted a genome-wide association study to identify runs of homozygosity (ROH), and copy number variation (CNV) using the novel Illumina PorcineSNP50 BeadChip array in Bama Xiang pigs and annotated the related genes in significant single nucleotide polymorphisms and common copy number variation region (CCNVR). We calculated the ROH-based genomic inbreeding coefficients (F_{ROH}) and the Spearman coefficient between F_{ROH} and reproduction traits. We completed a mixed linear model association analysis to identify the effect of high-frequency copy number variation (HCNVR; over 5%) on Bama Xiang pig reproductive traits using TASSEL software. Across eight chromosomes, we identified 29 significant single nucleotide polymorphisms, and 12 genes were considered important candidates for litter-size traits based on their vital roles in sperm structure, spermatogenesis, sperm function, ovarian or follicular function, and male/female infertility. We identified 9,322 ROHs; the litter-size traits had a significant negative correlation to F_{ROH}. A total of 3,317 CNVs, 24 CCNVR, and 50 HCNVR were identified using cnvPartition and PennCNV. Eleven genes related to reproduction were identified in CCNVRs, including seven genes related to the testis and sperm function in CCNVR1 (chr1 from 311585283 to 315307620). Two candidate genes (*NEURL1* and *SH3PXD2A*) related to reproduction traits were identified in HCNVR34. The result suggests that these genes may improve the litter size of Bama Xiang by marker-assisted selection. However, attention should be paid to deter inbreeding in Bama Xiang pigs to conserve their genetic diversity.

Keywords: GWAS, ROH, CNV, litter size, teat number, candidate genes, Bama Xiang pig

INTRODUCTION

Bama Xiang pigs, an indigenous Chinese pig breed, are famous for their excellent meat quality and early maturation (1). Owing to their small size, Bama Xiang pigs are easy to handle. Further, Bama Xiang pigs have anatomical and physiological traits similar to humans. Therefore, these pigs could be used in human medical research, such as hypertrophic scarring and diabetes (2, 3). However, genetic diversity in Bama Xiang pigs is declining due to historical inbreeding. Runs of homozygosity (ROH) arise when the same haplotypes are inherited from parents (4), especially in inbreeding Bama Xiang pigs. Bama Xiang boars sexually mature in 76 days and often mate with their mothers resulting in inbreeding (5). The increased inbreeding and declining genetic diversity may hamper the sustainable production of the Bama Xiang pig. Moreover, replacing traditional breeding with intensive pig farming has been beneficial to Bama Xiang pig breeding. The ROH analysis and evaluation of inbreeding rates are important for conserving Bama Xiang pig resources.

Copy number variations (CNVs) are a subtype of genomic structural variation ranging from 50 bp to several Mb in length. The copy number variation region (CNVR) is the area adjacent to the copy number with overlapping regions (6). The CNVR, owing to its length, has a higher probability of changing gene structure and gene dosage and is known to affect several traits in pigs. Qiu et al. demonstrated that nine CNVRs were associated with average daily gain and days to 100 kg in Duroc (7). Zheng et al. (8) reported that the copy number of the *AHR* had a positive effect on reproduction traits. The copy number variation in *GPER1* might be related to the litter size in the Large White pig breed (9). Bovo et al. (10) showed that the CNVR in *MSRB3* may be associated with the ear size. Thus, analyzing CNVR function has become an important part of porcine genetics and breeding.

Single nucleotide polymorphism (SNP) arrays can genotype hundreds of thousands of SNPs distributed throughout the genome (11). Based on the density, the porcine SNP arrays were split into 50K (12), 60K (13), and 80K (14). The beadchips have been widely used for genomic selection (15), selection signature research (14), and genome-wide association studies (16). The array can be used to complete population genetics research (17) and CNV detection (18).

Litter size and teat number, which are the base index, are typically associated with economic benefits and production ability in the porcine industry. While the teat number has a medium level of heritability, the litter-size heritability is low. During breeding, the teat number trait increases with an increase in the litter-size trait (19, 20). Litter size represents productivity levels per sow per year in pigs (21). Teat number is a proxy for lactation ability and, thus, is related to piglet mortality rates (22). Numerous researchers have focused on identifying SNPs, quantitative trait loci (QTLs), and candidate genes associated with litter size and teat number (23–25). However, the studies of quantitative traits based on CNVs have rarely been completed, especially on litter size and teat number. Therefore, a study on CNV associated with litter size and teat number in pigs is required.

In this study, we completed the genome-wide association study (GWAS) and the ROH analysis to evaluate the genomic inbreeding coefficients (F_{ROH}) of Bama Xiang pigs using SNP array data. We also performed CNV detection and CNVR-based association analysis of litter size and teat number of Bama Xiang pigs. We identified the genes in significant SNPs, ROH, and CNVR and provided the candidate genes associated with litter size and the teat number of Bama Xiang pigs. The information on the Bama Xiang pig can provide the development of the molecular mechanisms of litter size and teat number. Understanding the genetic basis of litter size and teat number in Bama Xiang pigs should help us improve their reproductive capacity.

MATERIALS AND METHODS

Animal and Phenotype Data

This study collected ear tissues of 403 Bama Xiang sows from the Agriculture and Animal Husbandry Co. Ltd. (Guangxi, China) for genomic DNA extraction. Parity and teat number data from 297 sows (2,199 dens) were also obtained to calculate phenotype statistics of 14 litter-size traits and five teat-number traits (**Supplementary Table S1**).

The DNA Subjects and Genotyping

Using a tissue DNA isolation mini kit (Vazyme, Nanjing, Jiangsu, China, Cat. #DC112-01), the genomic DNA was isolated from the ear tissue. Porcine SNP50 BeadChip (Illumina, Inc.) containing 51,315 SNPs was used to genotype the genomic DNA. The raw data were called using the GenomeStudio 2.0 software. Single nucleotide polymorphism arrays with call rates less than 0.9 or with minor allele frequencies of <0.05 were removed in PLINK, version 1.90. Single nucleotide polymorphism arrays without location information were also deleted. After quality control, 403 sows and 24,123 autosomal SNPs were used for analysis. Beagle, version 5.0, was used to impute missing alleles.

Genome-Wide Association Studies

For each trait, we implemented GWAS using a univariate linear mixed model in GEMMA. The GWAS model was as follows (26):

$$\mathbf{y} = \mathbf{W}\alpha + \mathbf{x}\beta + \mathbf{u} + \boldsymbol{\varepsilon}; \mathbf{u} \sim \mathrm{MVNn}\left(0, \lambda\tau^{-1}\mathbf{K}\right), \boldsymbol{\varepsilon} \sim \mathrm{MVNn}\left(0, \tau^{-1}\mathbf{In}\right),$$

where $\mathbf{y}$ is an n-vector of reproduction traits, $\mathbf{W}$ is a matrix of fixed effects including three principal constituents and the parity effect, $\mathbf{x}$ is the SNP genotype, α and β are the corresponding coefficients, $\mathbf{u}$ is the random effect, $\boldsymbol{\varepsilon}$ is the random error, τ^{-1} is the variance of residual errors, λ is the ratio between the two variance components, $\mathbf{K}$ is the kinship matrix, $\mathbf{In}$ is the $n \times n$ identity matrix, and MVNn denotes the n-dimensional multivariate normal distribution. Thresholds for Bonferroni-adjusted genome-wide significance and suggestive significance were defined as $-\log_{10}(p) = 5.68$ (0.05/24,123) and $-\log_{10}(p) = 4.38$ (1/24,123), respectively (27–29). We defined the 200 kb regions upstream and downstream of the saliency marker as significant for identifying genes in the Biomart program (*Sus scrofa 10.2*).

The ROH Detection and ROH-Based Genomic Inbreeding Coefficients

The ROH was identified using detectRUNS packages with the following parameters: Window size was 15; the threshold was 0.05; the minimum number of homozygous/heterozygous SNP in the window was 30; the maximum number of SNP with opposite genotype was 1; the maximum number of missing genotypes was 1; the maximum gap between consecutive SNPs was 250,000; minimum length of the run was 1,000,000; and number of SNPs every kilo-basepairs was 1/100. The ROH length was classified into four classes: 0–6, 6–12, 12–24, and 24–48 Mb. The F_{ROH} for Bama Xiang pigs was estimated using the following formula (30):

$$F_{ROH} = L_{ROH}/L_{AUTO}, \quad (1)$$

where L_{ROH} is the total length of ROH on autosomes and L_{AUTO} is the total length of the autosomes.

The relationship between F_{ROH} and Bama Xiang pig reproduction traits was calculated using the Spearman coefficient. The SPSS 19.0 was used to calculate the general lines model.

The CNV Detection and CNVR Annotation

The CNVs were detected using cnvPartition 3.2.0 and PennCNV 1.0.5 software. As a plug-in software in GenomeStudio 2.0, cnvPartition detected different copy numbers using Gaussian distribution by Log*R* ratio and B allele frequency values. To increase the accuracy of CNVs, the cnvPartition was completed with the following parameters: Confidence threshold, 35; minimum homozygous region size, 1,000,000; and minimum probe count, 3. PennCNV used the Log*R* ratio and B allele frequency values to identify the copy number base on the hidden Markov model with parameters: numSNP, 3; length, 1 K. The CNVs were detected using cnvPartition, and PennCNV was merged into CNVRs using bedtools software. According to the types of CNVRs, we merged the same type of CNVR from two software into CCNVR. The CCNVR of Bama Xiang pigs was annotated using the BioMart program and David database. The Pig QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index) was used to annotate the traits related to the CCNVR.

The CNVR-Based Association Analysis

According to the CNVR frequency detected by cnvPartition and PennCNV, we defined the CNVR frequency over 5% as the HCNVR. To identify the effect of HCNVR on Bama Xiang pigs' reproductive traits, we completed a mixed linear model association analysis with the Q + K method using TASSEL software, where Q is the principal component matrix and K is the kinship matrix. The Q and K matrices were calculated by SNP array data using Tassel software. According to the Bonferroni thresholds of GWAS, we defined $p = 0.001$ (0.05/50) and $p = 0.02$ (1/50) as the significance thresholds and suggestive significance thresholds, respectively (27–29). The significance and suggestive significance HCNVRs were annotated using the BioMart program.

RESULT

The Significant SNPs and Genes From GWAS

We found that 29 SNPs exceeded the suggestive significance threshold in SSC 1, 3, 4, 5, 7, 14, 16, and 17. Seven litter-size traits (**Figure 1**) and four teat-number traits (**Supplementary Figure S1**; **Supplementary Table S2**) were covered, including average birth number (ABN, one SNP), birth number–second (BN2, two SNPs), birth number–fourth (BN4, one SNPs), birth number–fifth (BN5, 10 SNPs), birth number–sixth (BN6, four SNPs), birth number–ninth (BN9, two SNPs), the maximum number of births (MAXBN, three SNPs), teat numbers on the left side (LTN, one SNP), teat numbers on the right side (RTN, one SNP), the total number of teats (TTN, three SNPs), and minimum teat number (MINTN, one SNP). The SNP rs80960023 on SSC4 was significant in ABN. The SNP rs80967544 was a significant SNP in BN5 and MAXBN. The SNP rs81450533 on SSC14 was significant in LTN, TTN, and MINTN. Forty-seven genes were significantly associated with litter-size traits (**Supplementary Table S3**). Notably, 12 genes were identified as related to reproduction, including *CIB4, DRC1*, *HADHA*, *DMC1*, *DDX17*, *RAB23*, *HMGA1*, *SLC26A8*, *MAPK14*, *ARRDC4*, *CDH6*, and *BMP7*.

The ROH Detection and Genomic Inbreeding Coefficients Evaluation

We detected 9,322 ROHs, with an average length of 88.78 Mb, in 403 Bama Xiang pigs. The number of ROH per animal ranged from 1 to 136, with the average number of ROH being 23.14 ± 13.74 (mean ± SD). The number of ROH for 0–6, 6–12, 12–24, and 24–48 Mb was 8,369 (89.77%), 831 (8.91%), 110 (1.18%), and 12 (0.13%), respectively. The genomic inbreeding coefficients per sow ranged from 0.0013 to 0.24, with the average genomic inbreeding coefficient as 0.036 ± 0.024 (mean ± SD). A total of 12 litter-size traits had a significantly negative relation with F_{ROH} ($p < 0.05$), including ABN (−0.28), BN1 (−0.19), BN2 (−0.19), BN3 (−0.25), BN4 (−0.15), BN5 (−0.20), BN6 (−0.23), BN8 (−0.17), BN9 (−0.22), MAXBN (−0.22), and the minimum numbers of birth (MINBN) (−0.25) (**Supplementary Table S4**). The general linear model about ABN and F_{ROH} was ABN = 10.359–15.875 $*F_{ROH}$, with R^2 was 0.62, and $p < 0.05$.

The CNV Detection and Annotation

We identified 2,920 and 397 CNVs using cnvPartition and PennCNV software, respectively. After the merge, we removed the length of CNVRs over 4 Mb. Finally, we got 197 CNVRs ranging from 0.036 to 3.72 Mb, including 159 deletions, 10 duplications, and 28 mixed in cnvPartition (**Figure 2A**). We got 45 CNVRs ranging from 0.093 to 2.97 Mb, including 26 deletions, six duplications, and 13 mixed in PennCNV (**Figure 2B**). We merged

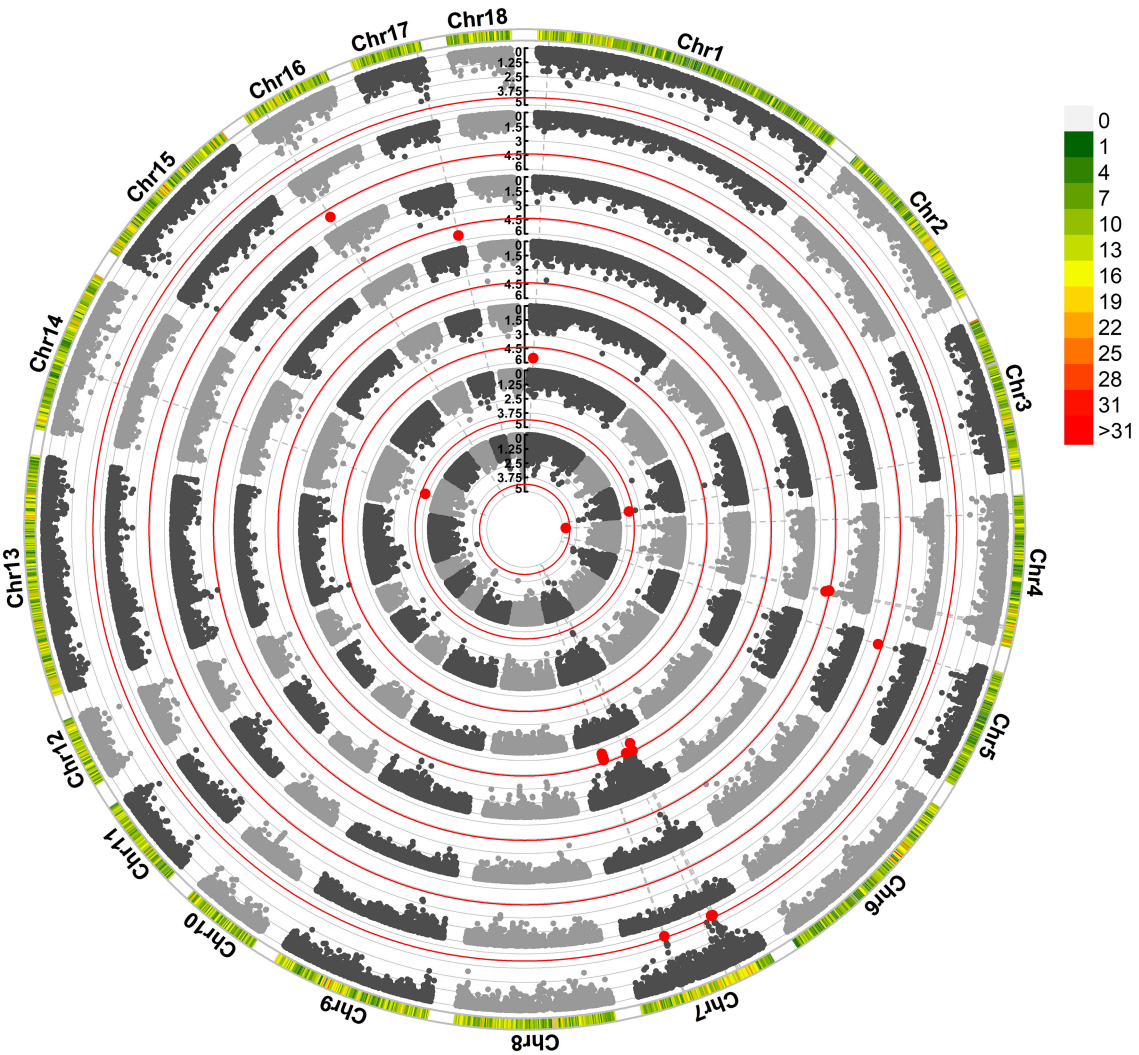


FIGURE 1 | Manhattan plots obtained in the GWAS of litter-size traits. The red line identifies the cut-off for suggestive significance. Red spots identify SNPs with suggestive significance. Traits, from the inner to outer lanes are average birth number (ABN), birth number–second (BN2), birth number–fourth (BN4), birth number–fifth (BN5), birth number–ninth (BN9), and maximum number of births.

the CNVRs from cnvPartition and PennCNV based on their type. Finally, we got 24 CCNVR, including 11 deletions, one duplication, and 12 mixed (**Figure 2C**; **Supplementary Table S5**). A total of 64 genes were located in CCNVRs. The genes were significantly enriched ($p < 0.05$) in the cilium movement and outer dynein arm assembly pathway. We located 637 QTLs in CCNVRs, including 26 average backfat thickness QTLs, 21 backfat at last rib QTLs, and 17 loin muscle area QTLs (**Supplementary Figure S2**).

The CNVRs Associated With Reproduction Traits

With cnvPartition and PennCNV software, we got 242 CNVRs in Bama Xiang pigs, including 50 HCNVR (**Figure 2D**; **Supplementary Table S6**). After execution of the association analysis, four HCNVRs were over the significance thresholds, and nine HCNVRs were over the suggestive significance thresholds. Interestingly, HCNVR10 (chr2 from 152455383 to 152748172) was over the suggestive significance thresholds in ABN and BN4, BN5, and MINBN (**Supplementary Table S7**). Additionally, HCNVR34 (chr14 from 124353052 to 124685417) was over the suggestive significance thresholds in ABN and MAXBN, the maximum number of teats (MAXTN), TTN, and LTN (**Supplementary Table S8**). We executed a *t*-test with reproduction traits between different CNVR types in HCNVR10 and HCNVR34 (**Supplementary Table S8**). The deletions were lower than normal in ABN, BN4, BN7, MINBN, and other reproductive traits such as HCNVR10. In ABN, MAXBN, MAXTN, TTN, and LTN, the phenotype of deletions was lower

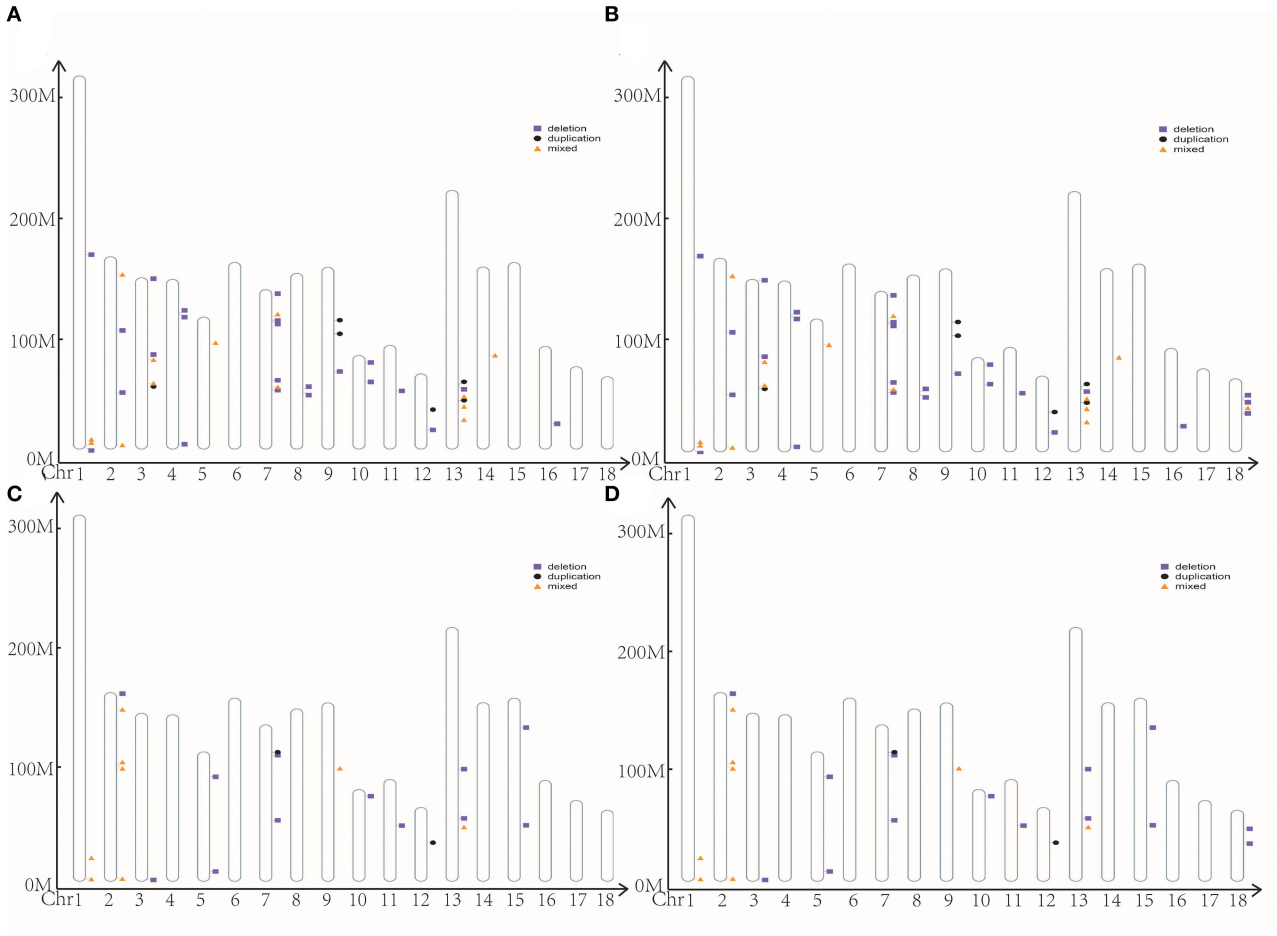


FIGURE 2 | The position of CNVR. **(A)** CNVR from cnvPartition; **(B)** CNVR from PennCNV; **(C)** the common CNVR; **(D)** the high frequency CNVR.

than normal in HCNVR34. After annotation, we got two genes (*NEURL1* and *SH3PXD2A*) related to reproduction traits located in HCNVR34.

DISCUSSION

The GWAS Reveals the Candidate Genes Related to Litter Size and Teat Number

We identified 18 genes by using GWAS. Of these, *MAPK14* was the only gene that was not associated with a specific sex; *MAPK14* promotes phosphorylation and is therefore involved in both male and female reproduction (31). It is critical for the heat-induced proliferation of spermatogenic cells. Moreover, *MAPK14* is mainly localized in granulosa and follicle cells in the ovaries, implying its importance during follicular development (32).

Most of the remaining genes were related to sperm structure, spermatogenesis, sperm function, and male infertility. Gene *CIB4*, which is strongly expressed in mouse and human testes, plays an essential role in the spermatid head. Gene *CIB4* knockout (KO) mice are sterile because haploid differentiation becomes impaired (33). Knocking out the *DRC1* gene in mice completely disorders the axoneme structure of sperm flagella, impairing sperm motility (34). Research in bulls shows that *HADHA* proteins are significantly more abundant in the immotile sperm of low-fertility males than in the sperm from high-fertility males (35). Gene *DMC1* plays a vital role in repairing DNA double-strand breaks, and disruption of this repair process is linked to male infertility (36). Gene *DDX17* increases during the transition from spermatocytes to sperm (37). Research in a Duroc × Erhualian F2 population identified *RAB23* as a candidate gene for pubertal reproductive failure (38). In humans, *HMGA1* is a stage-specific marker gene for germ cells, and in mice, it is essential for sperm production (39). Gene *SLC26A8* is a sperm-specific member of the *SLC26* family, and its heterozygous missense mutations are highly associated (power of >95%) with asthenozoospermia (40). Gene *ARRDC*4 mediates extracellular vesicle biogenesis, which appears to be required for sperm function, given that *ARRDC4* KO mice have

impaired sperm (41). Two of the 18 genes were related to ovaries, follicles, and female infertility. Gen *CDH6* regulates endometrial adhesion and implantation. Studies show that the *CDH6* gene is dysregulated in the endometrium of women with infertility (42). Gene *BMP7* regulates steroidogenesis, granulosa cell states, and follicular development. A study on Yorkshire pigs found that it is a candidate gene for litter size (43, 44).

The ROHs Reveal the F_{ROH} Effect on the Litter Size

In this study, we successfully identified 9,322 ROHs. Similar to Laiwu pig (45), Diannan xiaoer pig (46) and Large White (47), the majority ROHs in Bama Xiang pigs were short segments. The significant Spearman correlation coefficient of F_{ROH} and litter size ranged from −0.14 to −0.28. No significant correlation was noticed between the F_{ROH} and teat number. Our result, same as proposed by Tao et al. (48), illustrated that the inbreeding had a significant effect on the litter size but had no effect on the teat number in Bama Xiang pigs. According to the general linear model, the ABN decreased to 0.16 when the F_{ROH} was increased to 0.01. Thus, the average number of births decreased to 0.57 ± 0.38 in Bama Xiang pig. We established that decreased F_{ROH} improves the litter size in Bama Xiang pigs.

The CNV Reveals the Candidate Genes Related to Litter Size and Teat Number

We identified 3,317 CNVs, 242 CNVRs, and 13 CCNVRs in Bama Xiang pigs. A total of 11 genes related to reproduction, *TUBB4B* (49), *STPG3* (50), *SNRPA1* (51), *NRARP* (52), *NEUROD4* (53), *MGAT1* (54), *LRGUK* (55), *LCN8* (56), *LCN12* (57), *LCN10* (58), and *CCDC183* (58), were identified in CCNVR. Bama Xiang pigs are famous for their excellent meat quality and early maturation; however, their lower growth rate and higher backfat content make them less desirable. Thus, the result that the top 10 QTLs in CCNVRs are major related to backfat, meat quality, growth, and teat number is reasonable. However, reproduction-related QTLs were not found in CCNVR1. After executing association analysis, we discovered that HCNVR10 and HCNVR34 were associated with the reproduction traits in Bama Xiang pigs, including litter size and teat number. The phenotype of deletions of HCNVR10 and HCNVR34 were lower than normal in litter size and teat number. *NEURL1* and *SH3PXD2A* were located in HCNVR34. However, *NEURL1* and *SH3PXD2A*, possibly related to reproduction traits, have not been reported. The HCNVR34 (chr14 from 124353052 to 124685417) contained *NEURL1* and *SH3PXD2A* gene may be the candidate CNVR and candidate genes, which is related to litter size and teat number in Bama Xiang pigs.

The Genes Related to Testis and Sperm

In GWAS, we found that nine genes were related to testis and sperm, including *CIB4*, *DRC1*, *HADHA*, *DMC1*, *DDX17*, *RAB23*, *HMGA1*, *SLC26A8*, and *ARRDC4*. In CNV detection, we found that *TUBB4B*, *STPG3*, *NRARP*, *LCN8*, *LCN12*, *LCN10*, and *CCDC183* were related to the testis and sperm. This result indicated that the sperm of Bama Xiang pigs were different than those of other pigs. Our previous study found that the sperm from Duroc improved the litter size of Bama Xiang pig sows compared to the sperm from Bama Xiang boars (59). We speculate that the reduced litter size in Bama Xiang is due to inbreeding; however, the quality of sperm should be studied further to ascertain the cause of the reduced litter size. Several genes related to testis and sperm were identified from litter-size traits, illustrating that the marks play an important role in Bama Xiang pigs. Bama Xiang boars' sperm quality could be enhanced to improve the litter size of Bama Xiang sows. However, inbreeding in parent and offspring should be deterred to improve the fertility of Bama Xiang pigs.

CONCLUSIONS

In this study, we executed the GWAS, ROH analysis, and CNV detection using porcine 50K Beadchip in Bama Xiang pigs. A total of 29 candidate SNPs for seven litter-size traits and four teat-number traits were identified in Bama Xiang pigs using GWAS. Twelve candidate genes were identified in litter-size traits. A total of 9,322 ROHs were found, and the litter-size traits had a significant negative correlation with F_{ROH}. A total of 3,317 CNVs were identified, of which 11 genes may be the candidates for reproduction traits. Sixteen genes related to the testis and sperm function were identified. Our results confirm that by using marker-assisted selection on 16 genes, we can improve the litter size of Bama Xiang boars.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.6084/m9.figshare.19564270.v1.

ETHICS STATEMENT

The animal study was reviewed and approved by Guangxi University.

AUTHOR CONTRIBUTIONS

JM and YL designed the study, conducted the experiments, and drafted original manuscripts. SZ, LF, WQ, and XC conducted parts of experiments. BX, BC, GL, and JL revised the manuscripts. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the National Modern Agricultural Industrial Technology System (nycytxgxcxtd-15-01), the National Natural Science Foundation of China (81860150), and the Science and Technology Major Project of Guangxi (Guike-AA17292002).

ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China for funding this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.892815/full#supplementary-material

Supplementary Figure S1 | The Manhattan plots obtained in the GWAS with teat number traits. The red line identifies the cut-off for suggestive significance. Red spots identify SNPs with suggestive significance. Traits, from the inner to outer lanes are the teat numbers on the left side (LTN), the teat numbers on the right side (RTN), total number of teats (TTN), and the minimum number of teats (MINTN).

Supplementary Figure S2 | The top 10 QTLs in common CNVRs.

REFERENCES

1. Yang Y, Adeola AC, Xie HB, Zhang YP. Genomic and transcriptomic analyses reveal selection of genes for puberty in Bama Xiang pigs. *Zool Res.* (2018) 39:424–30. doi: 10.24272/j.issn.2095-8137.2018.068
2. Ning X, Yang K, Shi W, Xu C. Comparison of hypertrophic scarring on a red Duroc pig and a Guangxi Mini Bama pig. *Scars Burn Heal.* (2020) 6:2059513120930903. doi: 10.1177/2059513120930903
3. Zhang L, Huang Y, Wang M, Guo Y, Liang J, Yang X, et al. Development and genome sequencing of a laboratory-inbred miniature pig facilitates study of human diabetic disease. *iScience.* (2019) 19:162–76. doi: 10.1016/j.isci.2019.07.025
4. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* (2018) 19:220–34. doi: 10.1038/nrg.2017.109
5. Wang LY, Lin AG, Wang LX, Li K, Yang GS, He RG, et al. *Animal Genetic Resources in China Pigs*. Beijing: China Agriculture Press (2011), p. 237.
6. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* (2006) 444:444–54. doi: 10.1038/nature05329
7. Qiu Y, Ding R, Zhuang Z, Wu J, Yang M, Zhou S, et al. Genome-wide detection of CNV regions and their potential association with growth and fatness traits in Duroc pigs. *BMC Genomics.* (2021) 22:332. doi: 10.1186/s12864-021-07654-7
8. Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, et al. CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. *J Anim Sci Biotechnol.* (2020) 11:42. doi: 10.1186/s40104-020-00442-5
9. Wang Y, Zhang T, Wang C. Detection and analysis of genome-wide copy number variation in the pig genome using an 80 K SNP Beadchip. *J Anim Breed Genet.* (2020) 137:166–76. doi: 10.1111/jbg.12435
10. Bovo S, Ribani A, Muñoz M, Alves E, Araujo JP, Bozzi R, et al. Genome-wide detection of copy number variants in European autochthonous and commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. *Anim Genet.* (2020) 51:541–56. doi: 10.1111/age.12954
11. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE.* (2009) 4:e6524. doi: 10.1371/journal.pone.0006524
12. Zhang Z, Zhang Z, Oyelami FO, Sun H, Xu Z, Ma P, et al. Identification of genes related to intramuscular fat independent of backfat thickness in Duroc pigs using single-step genome-wide association. *Anim Genet.* (2021) 52:108–13. doi: 10.1111/age.13012
13. Salek Ardestani S, Jafarikia M, Sargolzaei M, Sullivan B, Miar Y. Genomic prediction of average daily gain, back-fat thickness, and loin muscle depth using different genomic tools in Canadian swine populations. *Front Genet.* (2021) 12:665344. doi: 10.3389/fgene.2021.665344
14. Wang X, Zhang H, Huang M, Tang J, Yang L, Yu Z, et al. Whole-genome SNP markers reveal conservation status, signatures of selection, and introgression in Chinese Laiwu pigs. *Evol Appl.* (2020) 14:383–98. doi: 10.1111/eva.13124
15. Melnikova E, Kabanov A, Nikitin S, Somova M, Kharitonov S, Otradnov P, et al. Application of genomic data for reliability improvement of pig breeding value estimates. *Animals*. (2021) 11:1557. doi: 10.3390/ani11061557
16. Xue Y, Li C, Duan D, Wang M, Han X, Wang K, et al. Genome-wide association studies for growth-related traits in a crossbreed pig population. *Anim Genet.* (2021) 52:217–22. doi: 10.1111/age.13032
17. Valluzzi C, Rando A, Macciotta NPP, Gaspa G, Di Gregorio P. The Nero Lucano pig breed: recovery and variability. *Animals.* (2021) 11:1331. doi: 10.3390/ani11051331
18. Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, et al. A genome-wide association study of copy number variations with umbilical hernia in swine. *Anim Genet.* (2016) 47:298–305. doi: 10.1111/age.12402
19. Martínez-Giner M, Noguera JL, Balcells I, Alves E, Varona L, Pena RN. Expression study on the porcine PTHLH gene and its relationship with sow teat number. *J Anim Breed Genet.* (2011) 128:344–53. doi: 10.1111/j.1439-0388.2011.00925.x
20. Merks JW, Mathur PK, Knol EF. New phenotypes for new breeding goals in pigs. *Animal.* (2012) 6:535–43. doi: 10.1017/S1751731111002266
21. Leman AD. Optimizing farrowing rate and litter size and minimizing nonproductive sow days. *Vet Clin North Am Food Anim Pract.* (1992) 8:609–21. doi: 10.1016/S0749-0720(15)30707-6
22. Chalkias H, Jonas E, Andersson LS, Jacobson M, de Koning DJ, Lundeheim N, et al. Identification of novel candidate genes for the inverted teat defect in sows using a genome-wide marker panel. *J Appl Genet.* (2017) 58:249–59. doi: 10.1007/s13353-016-0382-1
23. Li Y, Pu L, Shi L, Gao H, Zhang P, Wang L, et al. Revealing new candidate genes for teat number relevant traits in Duroc pigs using genome-wide association studies. *Animals (Basel).* (2021) 11:806. doi: 10.3390/ani11030806
24. Bovo S, Ballan M, Schiavo G, Ribani A, Tinarelli S, Utzeri VJ, et al. Single-marker and haplotype-based genome-wide association studies for the number of teats in two heavy pig breeds. *Anim Genet.* (2021) 52:440–50. doi: 10.1111/age.13095
25. Vashi Y, Magotra A, Kalita D, Banik S, Sahoo NR, Gupta SK, et al. Evaluation of candidate genes related to litter traits in Indian pig breeds. *Reprod Domest Anim.* (2021) 56:577–85. doi: 10.1111/rda.13895
26. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* (2012) 44:821–4. doi: 10.1038/ng.2310
27. Singh A, Kumar A, Gondro C, Pandey AK, Dutt T, Mishra BP. Genome wide scan to identify potential genomic regions associated with milk protein and minerals in vrindavani cattle. *Front Vet Sci.* (2022) 9:760364. doi: 10.3389/fvets.2022.760364
28. Zhou S, Ding R, Zhuang Z, Zeng H, Wen S, Ruan D, et al. Genome-wide association analysis reveals genetic loci and candidate genes for chest, abdominal, and waist circumferences in two duroc pig populations. *Front Vet Sci.* (2022) 8:807003. doi: 10.3389/fvets.2021.807003
29. Huang J, Wang C, Ouyang J, Tang H, Zheng S, Xiong Y, et al. Identification of key candidate genes for beak length phenotype by whole-genome resequencing in geese. *Front Vet Sci.* (2022) 9:847481. doi: 10.3389/fvets.2022.847481

30. Hidalgo J, Cesarani A, Garcia A, Sumreddee P, Larios N, Mancin E, et al. Genetic background and inbreeding depression in Romosinuano cattle breed in Mexico. *Animals.* (2021) 11:321. doi: 10.3390/ani11020321
31. Zou L, Cheng G, Xu C, Liu H, Wang Y, Li N, et al. The role of miR-128-3p through MAPK14 activation in the apoptosis of GC2 spermatocyte cell line following heat stress. *Andrology.* (2021) 9:665–72. doi: 10.1111/andr.12923
32. Hu S, Rao M, Lei H, Wu Y, Wang Y, Ke D, et al. Expression patterns of p38αMAPK during follicular development in the ovaries of neonatal rats. *Acta Histochem.* (2017) 119:538–42. doi: 10.1016/j.acthis.2017.05.007
33. Xu Z, Miyata H, Kaneda Y, Castaneda JM, Lu Y, Morohoshi A, et al. CIB4 is essential for the haploid phase of spermatogenesis in mice. *Biol Reprod.* (2020) 103:235–43. doi: 10.1093/biolre/ioaa059
34. Zhang J, He X, Wu H, Zhang X, Yang S, Liu C, et al. Loss of DRC1 function leads to multiple morphological abnormalities of the sperm flagella and male infertility in human and mouse. *Hum Mol Genet.* (2021) 30:1996–2011. doi: 10.1093/hmg/ddab171
35. D'Amours O, Calvo É, Bourassa S, Vincent P, Blondin P, Sullivan R. Proteomic markers of low and high fertility bovine spermatozoa separated by Percoll gradient. *Mol Reprod Dev.* (2019) 86:999–1012. doi: 10.1002/mrd.23174
36. Shang Y, Huang T, Liu H, Liu Y, Liang H, Yu X, et al. MEIOK21: a new component of meiotic recombination bridges required for spermatogenesis. *Nucleic Acids Res.* (2020) 48:6624–39. doi: 10.1093/nar/gkaa406
37. Pandey A, Yadav SK, Vishvkarma R, Singh B, Maikhuri JP, Rajender S, et al. The dynamics of gene expression during and post meiosis sets the sperm agenda. *Mol Reprod Dev.* (2019) 86:1921–39. doi: 10.1002/mrd.23278
38. Xin WS, Zhang F, Yan GR, Xu WW, Xiao SJ, Zhang ZY, et al. A whole genome sequence association study for puberty in a large Duroc × Erhualian F2 population. *Anim Genet.* (2018) 49:29–35. doi: 10.1111/age.12623
39. Wang M, Liu X, Chang G, Chen Y, An G, Yan L, et al. Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell.* (2018) 23:599–614.e4. doi: 10.1016/j.stem.2018.08.007
40. Dirami T, Rode B, Jollivet M, Da Silva N, Escalier D, Gaitch N, et al. Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am J Hum Genet.* (2013) 92:760–6. doi: 10.1016/j.ajhg.2013.03.016
41. Foot NJ, Gonzalez MB, Gembus K, Fonseka P, Sandow JJ, Nguyen TT, et al. Arrdc4-dependent extracellular vesicle biogenesis is required for sperm maturation. *J Extracell Vesicles.* (2021) 10:e12113. doi: 10.1002/jev2.12113
42. Zhou W, Santos L, Dimitriadis E. Characterization of the role for cadherin 6 in the regulation of human endometrial receptivity. *Reprod Biol Endocrinol.* (2020) 18:66. doi: 10.1186/s12958-020-00624-w
43. Du X, Yin H, Pan Z, Wu W, Shang P, Chamba Y, et al. BMP7 is a candidate gene for reproductive traits in Yorkshire sows. *Anim Reprod Sci.* (2020) 221:106598. doi: 10.1016/j.anireprosci.2020.106598
44. Li X, Ye J, Han X, Qiao R, Li X, Lv G, et al. Whole-genome sequencing identifies potential candidate genes for reproductive traits in pigs. *Genomics.* (2020) 112:199–206. doi: 10.1016/j.ygeno.2019.01.014
45. Fang Y, Hao X, Xu Z, Sun H, Zhao Q, Cao R, et al. Genome-wide detection of runs of homozygosity in Laiwu pigs revealed by sequencing data. *Front Genet.* (2021) 12:629966. doi: 10.3389/fgene.2021.629966
46. Wu F, Sun H, Lu S, Gou X, Yan D, Xu Z, et al. Genetic diversity and selection signatures within Diannan small-ear pigs revealed by next-generation sequencing. *Front Genet.* (2020) 11:733. doi: 10.3389/fgene.2020.00733
47. Shi L, Wang L, Liu J, Deng T, Yan H, Zhang L, et al. Estimation of inbreeding and identification of regions under heavy selection based on runs of homozygosity in a Large White pig population. *J Anim Sci Biotechnol.* (2020) 11:46. doi: 10.1186/s40104-020-00447-0
48. Tao L, He X, Wang X, Di R, Chu M. Litter size of sheep (*Ovis aries*): inbreeding depression and homozygous regions. *Genes.* (2021) 12:109. doi: 10.3390/genes12010109
49. Sarkar H, Arya S, Rai U, Majumdar SS. A study of differential expression of testicular genes in various reproductive phases of *Hemidactylus flaviviridis* (Wall Lizard) to derive their association with onset of spermatogenesis and its relevance to mammals. *PLoS ONE.* (2016) 11:e0151150. doi: 10.1371/journal.pone.0151150
50. Park S, Shimada K, Fujihara Y, Xu Z, Shimada K, Larasati T, et al. CRISPR/Cas9-mediated genome-edited mice reveal 10 testis-enriched genes are dispensable for male fecundity. *Biol Reprod.* (2020) 103:195–204. doi: 10.1093/biolre/ioaa084
51. Wu H, Sun L, Wen Y, Liu Y, Yu J, Mao F, et al. Major spliceosome defects cause male infertility and are associated with nonobstructive azoospermia in humans. *Proc Natl Acad Sci USA.* (2016) 113:4134–9. doi: 10.1073/pnas.1513682113
52. Murta D, Batista M, Silva E, Trindade A, Henrique D, Duarte A, et al. Dynamics of Notch pathway expression during mouse testis post-natal development and along the spermatogenic cycle. *PLoS ONE.* (2013) 8:e72767. doi: 10.1371/journal.pone.0072767
53. Kottmann JS, Jørgensen MGP, Bertolini F, Loh A, Tomkiewicz J. Differential impacts of carp and salmon pituitary extracts on induced oogenesis, egg quality, molecular ontogeny and embryonic developmental competence in European eel. *PLoS ONE.* (2020) 15:e0235617. doi: 10.1371/journal.pone.0235617
54. Biswas B, Batista F, Akintayo A, Aguilan J, Stanley P. Transgenic rescue of spermatogenesis in males with Mgat1 deleted in germ cells. *Front Cell Dev Biol.* (2020) 8:212. doi: 10.3389/fcell.2020.00212
55. Okuda H, DeBoer K, O'Connor AE, Merriner DJ, Jamsai D, O'Bryan MK. LRGUK1 is part of a multiprotein complex required for manchette function and male fertility. *FASEB J.* (2017) 31:1141–52. doi: 10.1096/fj.201600909R
56. Wen Z, Liu D, Zhu H, Sun X, Xiao Y, Lin Z, et al. Deficiency for Lcn8 causes epididymal sperm maturation defects in mice. *Biochem Biophys Res Commun.* (2021) 548:7–13. doi: 10.1016/j.bbrc.2021.02.052
57. Suzuki K, Lareyre JJ, Sánchez D, Gutierrez G, Araki Y, Matusik RJ, et al. Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. *Gene.* (2004) 339:49–59. doi: 10.1016/j.gene.2004.06.027
58. Vandenbrouck Y, Pineau C, Lane L. The Functionally Unannotated Proteome of Human male tissues: a shared resource to uncover new protein functions associated with reproductive biology. *J Proteome Res.* (2020) 19:4782–94. doi: 10.1021/acs.jproteome.0c00516
59. Mo JY, Gao JY, Zhao MW, Feng LL, Huang Y. Chen KY,et al. The effects of parity and boar on the litter size of Bama Xiang pigs and the sample size requirements for litter size statistics Chinese. *J Anim Sci.* (2020) 56:63–9. doi: 10.19556/j.0258-7033.20191230-03

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

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

frontiers | Frontiers in Veterinary Science

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fvets.2022.888207

OPEN ACCESS

EDITED BY
Aixin Liang,
Huazhong Agricultural University, China

REVIEWED BY
Tatiana Deniskova,
L.K. Ernst Federal Science Center for Animal Husbandry (RAS), Russia
Xiangpeng Yue,
Lanzhou University, China

*CORRESPONDENCE
Mingxing Chu
mxchu@263.net

†These authors have contributed equally to this work

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

RECEIVED 02 March 2022
ACCEPTED 05 July 2022
PUBLISHED 22 July 2022

CITATION
Du X, He X, Liu Q, Liu Q, Di R and Chu M (2022) Identification of photoperiod-induced specific miRNAs in the adrenal glands of Sunite sheep (*Ovis aries*). *Front. Vet. Sci.* 9:888207. doi: 10.3389/fvets.2022.888207

Identification of photoperiod-induced specific miRNAs in the adrenal glands of Sunite sheep (*Ovis aries*)

Xiaolong Du[1†], Xiaoyun He[1†], Qingqing Liu[2], Qiuyue Liu[1], Ran Di[1] and Mingxing Chu[1*]

[1]Key Laboratory of Animal Genetics, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, Institute of Animal Sciences of Chinese Academy of Agricultural Sciences, Beijing, China, [2]College of Animal Science and Technology, Anhui Agricultural University, Hefei, China

In seasonal estrus, it is well known that melatonin-regulated biorhythm plays a key role. Some studies indicate that the adrenal gland plays an important role in reproduction in mammals, but the molecular mechanism is not clear. This study used an artificially controlled light photoperiod model, combined with RNA-seq technology and bioinformatics analysis, to analyze the messenger RNA (mRNA) and microRNA (miRNA) of ewe (Sunite) adrenal glands under different photoperiod treatments. After identification, the key candidate genes *GRHL2*, *CENPF*, *FGF16* and *SLC25A30* that photoperiod affects reproduction were confirmed. The miRNAs (oar-miR-544-3p, oar-miR-411b-5p, oar-miR-376e-3p, oar-miR-376d, oar-miR-376b-3p, oar-miR-376a-3p) were specifically expressed in the adrenal gland. The candidate mRNA-miRNA pairs (e.g., *SLC25A30* coagulated by novel miRNA554, novel miRNA555 and novel miRNA559) may affect seasonal estrus. In summary, we constructed relation network of the mRNAs and miRNAs of sheep adrenal glands using RNA sequencing and bioinformatics analysis, thereby, providing a valuable genetic variation resource for sheep genome research, which will contribute to the study of complex traits in sheep.

KEYWORDS

sheep, adrenal gland, photoperiod, reproduction, RNA-seq

Introduction

Low fecundity is the greatest limiting factor in the modern mutton sheep industry. Many element affect fecundity, such as litter size, oestrus frequency, and embryo survival rate. Because the most part area of China is located in the temperate zone, many animals are in oestrus in autumn and winter and give birth in spring and summer to ensure the survival of their offspring. Biologists are therefore divided into long-photoperiod breeds (LPs) and short-photoperiod breeds (SPs) with seasonal variations (1, 2). For example, the Chinese Sunite sheep is a typical SP breeder, which is specifically represented as estrus from August to March of the next year and anestrus from April to July (3, 4).

Concerning reproduction, the unavoidable theme is the upstream control center, such as the hypothalamus and pituitary gland. Studies have shown that light stimulates

the paraventricular nucleus (PVN) and then affects the production of melatonin in the pineal gland, ultimately acting on the hypothalamus and affecting the reproduction of sheep through the hypothalamus-pituitary-gonad axis (HPG) (5–9). In addition, stimulating PVN also secretes corticotropin-releasing hormone (CRH), which then activates the pituitary gland to release corticotropin (ACTH). ACTH in turn stimulates the adrenal gland to secrete cortisol, which then provides negative feedback to the brain in the classic steady-state feedback loop to regulate hypothalamus-pituitary-adrenal gland axis (HPA) signals (10). However, the close relationship between HPA and HPG is more than that. When the fetus develops in the uterus, the hormone system that regulates the HPG and HPA axes plays an important role in the growth and development of its tissue. Fetal glucocorticoid concentrations increase in the third trimester of pregnancy, which is conducive to the modification of fetal key tissues or organs to promote fetal survival, including lung maturation and pulmonary surfactant production (11). In a study of rodents, exposure to glucocorticoids was found to closely affect fetal development during intrauterine development, such as gonadogenesis, the establishment of the HPG axis and the reproductive tract's morphogenesis (12, 13). In addition, the increase in the concentration of glucocorticoids interferes with the concentration of serum testosterone (T) because it inhibits T biosynthesis (14).

With the development of high-throughput sequencing, RNA sequencing (RNA-seq) is increasingly widely used in livestock. Including sheep (15) and cattle (16), this technique is used to obtain the expression profile information of mRNA and miRNA, which makes a great contribution to revealing some important traits and mining their candidate genes (17). Research on miRNAs has shown a trend of the great outbreak in recent years. Precursor miRNAs are well known to be transcribed mainly by RNA polymerase II, digested by dicer and then processed into mature miRNAs (18). Many studies have shown that miRNA plays a key role in regulating animal phenotype, such as affecting wool curvature (19), immunity to infectious diseases (20) and fat deposition (21). In this study, we mainly explored the key miRNAs affecting reproductive traits and adrenal-specific expression of miRNAs in sheep adrenal tissue under different photoperiod conditions. Finally, the relationship between differentially expressed mRNAs and miRNAs was predicted by bioinformatics software, and an interaction network is constructed, which is expected to mine effective information.

Abbreviations: mRNA, message RNA; miRNA, microRNA; HPA, The hypothalamic–pituitary–adrenal; LP, Long photoperiod; SP, Short photoperiod; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LP42, The LP lasts for 42 days; SP-LP42, The SP lasts for 42 days, and then convert it to LP for 42 days; SP42, The SP lasts for 42 days; *GRHL2*, grainy head-like protein 2 homologs; *CENPF*, centromere protein F; *FGF16*, Fibroblast growth factor 16; *SLC25A30*, solute carrier family 25, member 30.

Materials and methods

Ethics statement

The experimental animals involved in this study were carried out after being examined by the Animal Experimental Welfare Ethics Committee of the Institute of Animal Sciences of Chinese Academy of Agricultural Sciences (IAS-CAAS, Beijing, China). In addition, the review acceptance number is No. IAS2018-3, and all the experimental procedures are executed in accordance with the relevant guidelines and regulations formulated by the Ministry of Agriculture and Rural Affairs of the People's Republic of China.

Preparation of animals

Nine non-pregnant adult Sunite ewes (aged 2–3 years old; weight 30–40 kg), which were randomly selected from a farm in Bayan Nur City (40°75′ north latitude), Inner Mongolia Autonomous Region, China, were used for the study. The selected ewes were uniformly transported to a farm at the Tianjin Institute of Animal Sciences, Tianjin (39°13′ north latitude), China, and the follow-up experiments were carried out after a month of routine feeding to adapt to the local environment. The ovaries of these animals were removed by surgery, and an estrogen silicone tube (E_2, Sigma Chemical Co., St. Louis, MO, USA) was implanted subcutaneously in the neck of the sheep to maintain plasma estradiol levels of 3–5 pg/ml, as described previously (1, 22, 23). The ewe postoperative recovery lasted for 30 days before artificial light period control. The ewes were randomly divided into three groups: SP42 (short photoperiod for 42 days; $n = 3$), LP42 (long photoperiod for 42 days; $n = 3$) and SPLP42 (short photoperiod for 42 days followed by a long photoperiod for 42 days; $n = 3$). The conditions for the long photoperiod were 16 h of light per day and 8 h without light. The lighting duration setting for the short photoperiod was the opposite of the lighting duration setting for the long photoperiod. All sheep had *ad libitum* feeding and drinking in an enclosed climate control chamber with only artificial light sources.

Tissues acquisition, library construction and sequencing

Adrenal gland tissue from euthanized ewes was quickly preserved in liquid nitrogen. Then, the stored tissues were used

for RNA extraction with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions.

The mRNA library was constructed with 3 μg of high-quality RNA using the NEB Next Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, USA) according to the manufacturer's instructions. During this process, Ribo-Zero™ Gold Kits (TIANGEN, Beijing, China) were used to remove rRNA. In addition, we used the UNG enzyme to degrade the second strand of U-containing cDNA and performed polymerase chain reaction (PCR) amplification to obtain an RNA library. Then, a PE150 (paired-end 150 bp, PE150) sequencing approach for mRNAs was performed on a Hiseq X platform (Illumina, San Diego, CA, USA).

The miRNA library was built by a Small RNA Sample Pre kit (TIANGEN). We directly took total RNA as the starting sample, added adaptors at both ends of small RNA, and then reverse-transcribed the RNA to synthesize cDNA. After PCR amplification, the target DNA fragments were separated by polyacrylamide gel electrophoreses (PAGE), and the 140- to 160-bp ligation products were recovered to generate a cDNA library. In addition, an SE50 (single-end 50 bp, SE50) sequencing approach for miRNAs was performed on the Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA). All sequencing data were outsourced to Annoroad Gene Technology Co., Ltd. (Beijing, China).

Data processing and transcriptome assembly

Bcl2fastq (v2.17.1.14) was used to process the offline data and convert the original image file into row sequencing reads on-base calling, which were row reads. Using an in-house Perl script made by Annoroad Genentech Co., Ltd. (Beijing, China) to remove low-quality reads, reads with adaptor contamination and reads with a rate of N > 5%, the clean mRNA reads were acquired from the raw reads. We used the Ovis aries reference genome (Oar_v4.0) and the genome annotation file from Ensembl. Cleaned reads were then mapped to the reference genome using HiSAT2 (v2.0.5) (24), and StringTie (v1.3.2d) was used to assemble the transcripts (25). HiSAT2 was run with "-rna - strandness RF" and "-dta -t -p 4," String Tie with "-G ref.gtf -rf—1," and the other parameters were set as the default. In addition to the above steps, the following steps were added to obtain the clean miRNA reads. These steps include removing reads without a 3'adaptor and insert fragment, removing the reads containing consecutive A/T/G/C bases, and removing the reads with abnormal final length. To ensure the accuracy of the subsequent analysis, the clean reads of sRNA sequencing were mapped to the reference genome (Oar_v4.0) by the comparison analysis software Bowtie v1.1.2.

Classification notes of sRNA and identification of miRNA

We obtained the situation, in which the sequence matched different regions in each sample by mapping the clean reads to the Ovis aries sequence in the miRBase database (Release 21) (26). At the same time, the known miRNA can be identified. The clean reads that were not annotated as a known miRNA were compared with the ncRNA sequence in Rfam (13.0) (27) to realize the annotation of rRNA, tRNA, snRNA, snoRNA and other ncRNA. The RepeatMasker program was used to comment on different types of repeats for clean reads that were not annotated as known miRNAs and ncRNAs (28). After identifying the above sRNA types and then using the matching (100% positional overlap) results with the location information of exons and introns of the gene, the sRNA from mRNA will be annotated (29). For sRNA reads that did not match the above-known annotation type, the software miRDeep2 (30) was used to predict the new miRNA, and the sequence, expression and structure information of each new miRNA were analyzed. Different mature body sequences, precursor sequences and positions will be considered different new miRNAs.

Differential expression and functional enrichment analysis of miRNA and mRNAs

The transcripts per million (TPM) and the fragments per kilobase per million mapped reads (FPKM) values were calculated to represent miRNA and mRNA expression levels based on the read number (31). The difference analysis was carried out by the software DESeq (32). The screening conditions of differentially expressed miRNAs were pval $\leq$ 0.05, padj $\leq$0.05 and $\log_2|$ (fold change) $|\geq$1. The differential expression criteria of mRNA were $|\log_2 \text{Ratio}| \geq 1$ and $q < 0.05$. In addition, the target genes of known and novel differentially expressed miRNAs were predicted by miRanda and TargetScan software, respectively (33). The intersection of the prediction results of the two software programs was selected as the target gene of the miRNAs.

We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses based on the targeted genes of differentially expressed miRNAs (DEMs) and differentially expressed mRNAs (DEGs). The hypergeometric test method was applied to assess significantly enriched GO terms and KEGG pathways, and FDR <0.05 was considered to be significantly enriched.

Construction of integral miRNA–mRNA interaction networks

To further describe the interaction between miRNA and mRNA, we used miRanda and TargetScan software for prediction and took the intersection of the prediction results of the two software programs as the target genes of miRNA. According to the identified DEMs and differentially expressed target mRNAs of DEMs, a regulatory network diagram was drawn by Cytoscape software (34).

Data validation

Transcripts ($n = 4$) were randomly selected and the primers were designed by Primer 5.0 software. The miRNA primers were synthesized by Shanghai Sangon Biotech. The miRNA quantitative (q)PCR conditions were as follows: 95°C for 15 min, followed by 40 cycles of 94°C for 20 s and 60°C for 34 s (miRcute Plus miRNA qPCR Kit, TIANGEN). In addition, U6 small nuclear RNA (snRNA; for miRNA) were utilized as reference genes. The data obtained from the qPCR were evaluated using the $2^{-\Delta\Delta Ct}$ method.

Results

Overview of sequencing data in adrenal gland tissue

We obtained the global expression profile of mRNA and miRNA in the adrenal glands of Sunite ewes under different photoperiod treatments. The data of subsequent analysis, namely, clean reads, were based on the filtered original offline data, which had no adapter-polluted reads, and the base error rate was <0.1%. Overall, we obtained 1,029 million clean reads of mRNA and 244 million clean reads of miRNA. To obtain more accurate sequences and the accuracy of subsequent analysis, bioinformatics analysis software was used to map clean reads to the reference genome and compare the results statistically. The results are shown in Table 1. We found that the average comparison rate of miRNA was ∼50%, while the average comparison rate of mRNA was approximately 95%. In addition, 123 million perfect match reads of miRNA and 978 million perfect match reads of mRNA were obtained separately (Table 1; Supplementary Table 1).

Identification of mRNAs and miRNAs in adrenal gland tissue

A total of 18,947 mRNAs were identified (Supplementary Table 2) after mapping to the sheep genome. According to the *Ovis aries* gene annotation files in the related database, we counted the number and proportion of the unique alignment sequences on the three functional elements of the gene (exon, intron and intergenic). In general, if the annotation information of the species is more comprehensive, most of the sequences should be aligned to the exon region, but alternative splicing and noise expression will cause some sequences to come from the intron region and the intergenic region. The following figure shows the distribution of unique alignment sequences in various regions of the genome and chromosome. Our results suggested that many mRNAs were situated in the intergenic region (more than 45%), followed by the exon (approximately 33%) and intron (nearly 20%) regions (Figure 1A; Supplementary Table 3). Furthermore, the chromosome distribution of mRNAs showed that chromosome 3 and chromosome 1 expressed the most genes, accounting for 9.62 and 9.65%, respectively, followed by chromosome 2 (7.47%) and chromosome 5 (5.67%) (Figure 1B; Supplementary Table 4).

Regarding miRNAs, RNA-seq generated approximately 244 million clean reads after filtering and mapping. We classified and annotated all clean reads, including known miRNAs, ncRNAs (rRNA\ tRNA\ snRNA\ snoRNA\ other rfam), repeats, and small RNAs annotated by perfect matching with mRNA exons and introns and novel miRNAs. The detailed classification and specific statistics are shown in Table 2. The proportion of clean reads matching known miRNA maturity in each sample was more than 70%, followed by the proportion of clean reads matching with other ncRNA in the rfam database being more than 15%. In addition, we also counted the number of clean read types that matched the miRNA maturity, namely, unique clean reads. The statistical results are shown in Figure 2A. We found that the following four types, namely, exon+, other, rRNA and intron+, accounted for more than 10% of unique clean reads in the three experimental groups. Compared with the statistical results of total reads, the proportion of unique clean reads with known miRNA typeS decreased significantly, from ∼70 to 1% (Supplementary Table 6). The distribution of chromosomes is similar to the distribution of chromosomes of mRNA. The results of Figure 2B show that the X chromosome contains the most miRNA (12.25%), followed by chromosomes 3 (9.86%), 2 (8.73%) and 1 (7.89%) (Supplementary Table 4).

Furthermore, our main concern is the identification of known miRNAs and novel miRNAs. In total, we identified miRNAs ($n = 861$), including known miRNAs ($n = 151$) and novel miRNAs ($n = 710$). The numbers of total clean reads identified as known miRNAs and novel miRNAs in the three experimental group samples were 93,377,486 and 177,217, respectively, and the numbers of unique clean reads identified as known miRNAs and novel miRNAs in the samples were 18,691 and 1,054, respectively. Figure 3 shows an example of miRDeep2 output. The top of Figure 3 shows the scores assigned to each part of the miRNA and the total count. Different colors are used to represent different parts of the predicted hairpin secondary

TABLE 1 Summary of the mapping data from the adrenal gland tissues.

Sample	Total reads	Mapped reads	Mapping rate (%)	Total reads	Mapped reads	Mapping rate (%)
	miRNA			mRNA		
LP42-1	28587504	13967493	48.86	111648952	105998978	94.94
LP42-2	27102088	12510029	46.16	101268666	95551270	94.35
LP42-3	27778108	14175856	51.03	111612316	106337978	95.27
SP42-1	20353895	10420944	51.2	125490436	119791918	95.46
SP42-2	30012282	15368864	51.21	110469132	104921437	94.98
SP42-3	26406199	13258597	50.21	115155856	109685478	95.25
SPLP42-1	24951610	12495852	50.08	121476492	115054106	94.71
SPLP42-2	29122928	14768748	50.71	120023448	114183042	95.13
SPLP42-3	29856881	16144823	54.07	111722560	106093325	94.96
Total	244171495	123111206		1028867858	977617532	

LP42, and SP42 represent a long light period every day for 42 days and a short light period every day for 42 days. SP-LP42 means a short light period every day for 42 days, followed by a long light period for 42 days.

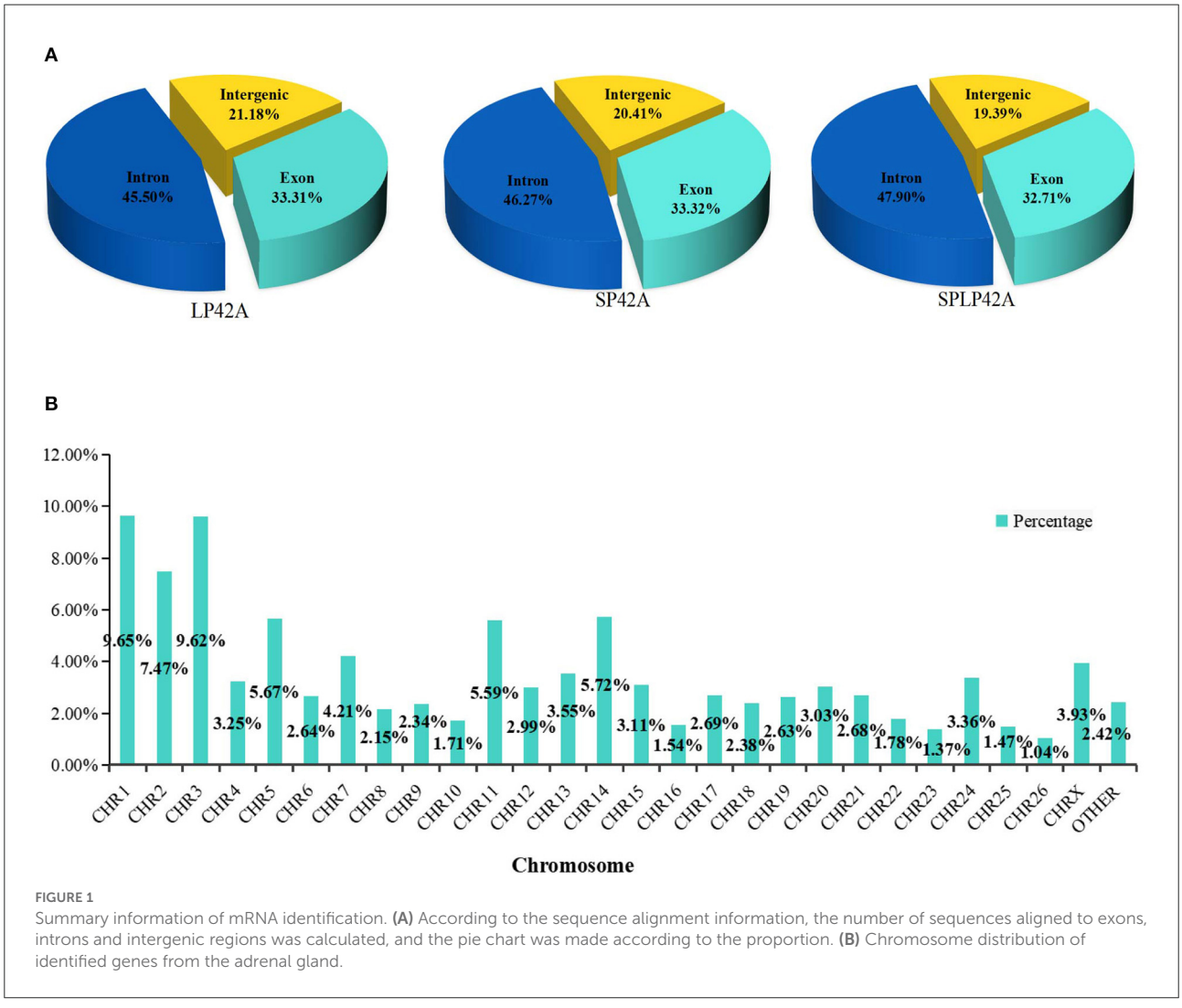


FIGURE 1
Summary information of mRNA identification. **(A)** According to the sequence alignment information, the number of sequences aligned to exons, introns and intergenic regions was calculated, and the pie chart was made according to the proportion. **(B)** Chromosome distribution of identified genes from the adrenal gland.

TABLE 2 Summary of the classification statistics of sRNA total reads mapped to reference genome.

Types	LP42	Proportion	SP42	Proportion	SPLP42	Proportion
Total	40653378	100%	39048405	100%	43409423	100%
Known miRNA	29903694	75.20%	29418632	73.61%	34055160	78.20%
rRNA	952232	2.18%	879015	2.37%	664044	1.50%
tRNA	522878	2.67%	947718	1.29%	495703	1.18%
snRNA	16647	0.03%	12853	0.04%	11207	0.02%
snoRNA	820367	2.06%	786458	2.02%	690074	1.59%
Other rfam	7321624	15.05%	5898519	17.92%	6508957	15.22%
Repeat	93962	0.20%	79954	0.23%	83080	0.19%
Exon:+	210897	0.53%	207450	0.52%	244056	0.57%
Exon:−	10167	0.03%	10336	0.03%	10320	0.02%
Intron:+	287885	0.70%	275741	0.71%	261042	0.61%
Intron:−	46007	0.20%	86753	0.12%	39145	0.09%
Novel miRNA	52460	0.12%	46912	0.13%	77845	0.18%
Other	414558	1.04%	398064	1.02%	268790	0.63%

The data of LP42, SP42 and SPLP42 are the total number of three biological repetitive samples, respectively. All the data and proportion calculations are based on mapped reads.

structure. All readings related to miRNA are also shown at the bottom of the figure. Therefore, about the detailed miRDeep2 output information of novel miRNAs, we described in detail the three biological repetitive samples of the three processing groups (LP42-1 $n = 115$, LP42-2 $n = 120$, LP42-3 $n = 119$, SP42-1 $n = 94$, SP42-2 $n = 138$, SP42-3 $n = 110$, SPLP42-1 $n = 117$, SPLP42-2 $n = 120$, SPLP42-3 $n = 121$), including the precursor secondary structure, count number and read sequence shown in the picture. Known miRNA is also described in detail (LP42-1 $n = 106$, LP42-2 $n = 106$, LP42-3 $n = 106$, SP42-1 $n = 104$, SP42-2 $n = 106$, SP42-3 $n = 105$, SPLP42-1 $n = 106$, SPLP42-2 $n = 106$, SPLP42-3 $n = 106$). The known miRNA specifically expressed in the adrenal gland was found by comparing with Zhang's Small Tail Han sheep hypothalamic miRNA transcriptome data (17), only six miRNAs (oar-miR-544-3p, oar-miR-411b-5p, oar-miR-376e-3p, oar-miR-376d, oar-miR-376b-3p, oar-miR-376a-3p) were specifically expressed in the adrenal gland, and one (oar-miR-1193-3p) was specifically expressed in the hypothalamus (Supplementary Tables 5, 6, miRDeep2 file).

The analysis of differentially expressed miRNAs and mRNAs in adrenal gland tissue

The numbers of DEGs and DEMs identified from the LP42 vs. SPLP42 comparison group were 144 and 48, respectively. Among these DEGs and DEMs, 45 and 10 genes were upregulated, and 99 and 38 genes were downregulated. In the same way, we counted the DEGs and DEMs of the SP42 vs. LP42 (DEG n = 454, DEM n = 36) and SP42 vs. SPLP42 (DEG $n = 506$, DEM $n = 55$) comparison groups. After analyzing the differences between all known miRNAs and novel miRNAs, we found that in known miRNAs, only the oar-miR-148a of the LP42 vs. SPLP42 comparison group was differentially expressed, so the graph shows only the statistics of novel miRNAs. Overall details are shown in Figure 4; Supplementary Table 7.

GO and KEGG enrichment analysis of DEGs and target gene of DEMs

To better understand the potential functions of the DEGs and target genes of DEMs, GO term and KEGG pathway analyses were performed. For DEGs, the three comparison groups we set up showed large differences. We found that the number of GO terms for biological processes significantly enriched in the three comparison groups was more than 50; in particular, the term number in the SP42 vs. SPLP42 comparison group was as high as 665. We selected only the top five terms in each group to draw the resulting chart. Second, we found a total of 10 terms in the three comparison groups regarding cellular components, so they were all drawn on the resulting graph. For the part of molecular function, except that we only found 4 significant terms in LP42 vs. SPLP42, the other two comparison groups chose the top five terms to draw the resulting map (Supplementary Table 9). Our main focus on BP terms included regulation of multicellular organismal process and regulation of the developmental process. The NAADP-sensitive calcium-release channel activity of the MF term also attracted our attention. The results of the KEGG pathway analysis are shown in Figure 5, and all the significantly enriched pathways found by the three comparison groups are plotted on the resulting map (Supplementary Table 10). The main pathways that attracted our

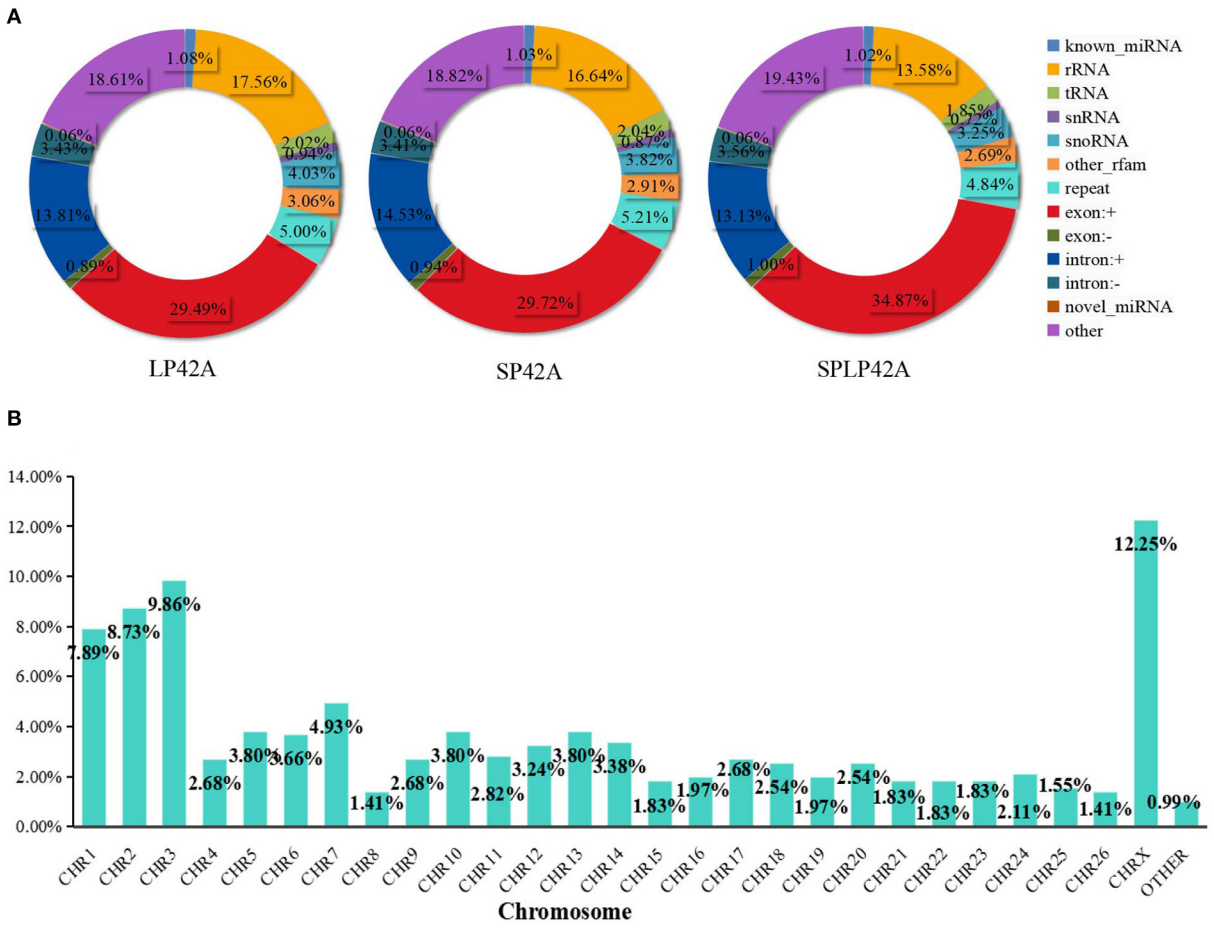


FIGURE 2
Summary information of sRNA identification. **(A)** The statistical results of unique clean reads from the adrenal gland in the three experimental groups. **(B)** Chromosome distribution of identified miRNA from the adrenal gland.

attention were cytokine-cytokine receptor interactions, the TNF signaling pathway, the Jak-STAT signaling pathway and the MAPK signaling pathway.

Regarding the target genes of miRNAs, LP42 vs. SPLP42 ($n = 119$), SP42 vs. LP42 ($n = 245$) and SP42 vs. SPLP42 ($n = 224$) differential target genes were found (Supplementary Table 8). The functional enrichment analysis of miRNA target genes was divided mainly into known miRNAs and novel miRNAs. The results are shown in Table 3. In the taxonomic list of known miRNAs, we found only that there was significant enrichment of GO terms in the LP42 vs. SPLP42 comparison group, except that the first five terms were selected in the BP, and all the significantly enriched terms about CC and MF are shown in the rest of the list. We are most interested in Wnt-activated receptor activity, Wnt-protein binding and cytokine binding. However, the results of GO enrichment analysis of novel miRNA target genes showed that only 13 terms in the MF part of the SP42 vs. SPLP42 comparison group were significantly enriched (Supplementary Table 11). Finally, we were also surprised to find that there was no significant enrichment of pathways in either known miRNAs classification or novel miRNAs classification in the three comparison groups in the KEGG enrichment analysis. However, there were some pathways with $p \leq 0.05$, not $Q \leq 0.05$ (Supplementary Table 12).

Analysis of integrated miRNA–mRNA co-expression network

To fully understand the potential roles of miRNAs, we built interactome networks using DEMs and their targets (DEGs). In total, 16 DE miRNAs (novel miRNAs) in SP42 vs. LP42 were predicted to target 67 genes. An mRNA–miRNA co-expression network was then constructed, where 1 DEG was targeted by 5 novel miRNAs. Regarding SP42 vs. SPLP42, 35 novel DE miRNAs were predicted to target 61 genes. The other

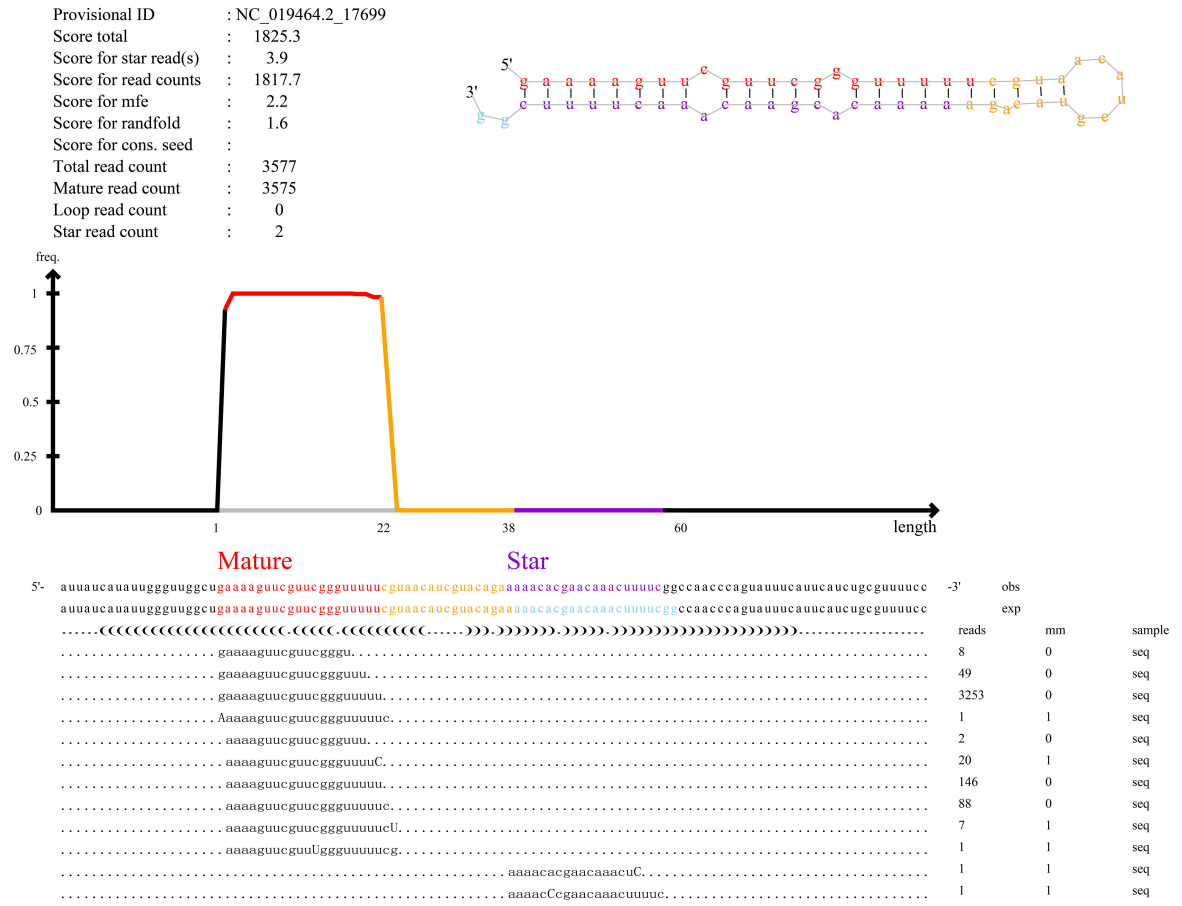


FIGURE 3
An example of the miRDeep2 output. The figure illustrates the output for the miRNA of novel_555. The upper part of Figure 3 shows the read count and the total count of each part. The predicted secondary structure of the hairpin is also depicted, with mature (red), star (purple) and circular (yellow) sequences highlighted in different colors. The bottom of the map shows the results of miRNA alignment with predicted precursor sequences on the genome (OBS line) and experimental sequences reported in miRBase (EXP line). For each sequence, the frequency and mismatch with the genome sequence (mm sequence) are given. Mismatches are also highlighted in uppercase letters.

mRNA–miRNA co-expression network was then constructed, where 4 DEGs were targeted by 16 novel miRNAs (Figure 6; Supplementary Tables 8, 13).

Data validation

To evaluate the accuracy of sequencing, qPCR was used to verify the RNA-seq data. The results showed that the expression pattern of miRNAs in sheep adrenal glands was similar to the expression pattern of sequencing (Figure 7), which proved the reliability of the data produced by RNA-seq.

Discussion

The transcriptome is space-time specific, but it is a powerful means to mine key information in the process of animal genetic development. The purpose of this study was to establish an OVX model and assist the light-controlled environment to identify the key sheep miRNAs and differentially expressed genes affected by light changes. Our previous studies successfully obtained the miRNAome and mRNAome information of the hypothalamus and pituitary through a similar experiment (17, 22). Therefore, we selected an adrenal transcriptional group for detailed analysis based on the close relationship between HPG and the HPA axis and successfully identified new adrenal-specific miRNAs and differentially expressed genes.

In this study, a total of 861 miRNA and 17,712 coding protein genes were identified in the three comparison groups. Among the miRNAs we identified, novel miRNAs accounted for approximately 82.45% of the total number of miRNAs, and known miRNAs accounted for only 17.54%. Among the 151 mature known miRNAs, we obtained only 106 stem-loop sequences. Both the 5 p arm and 3 p arm of 47 stem-loop sequences produced mature miRNA. Both arms of a miRNA

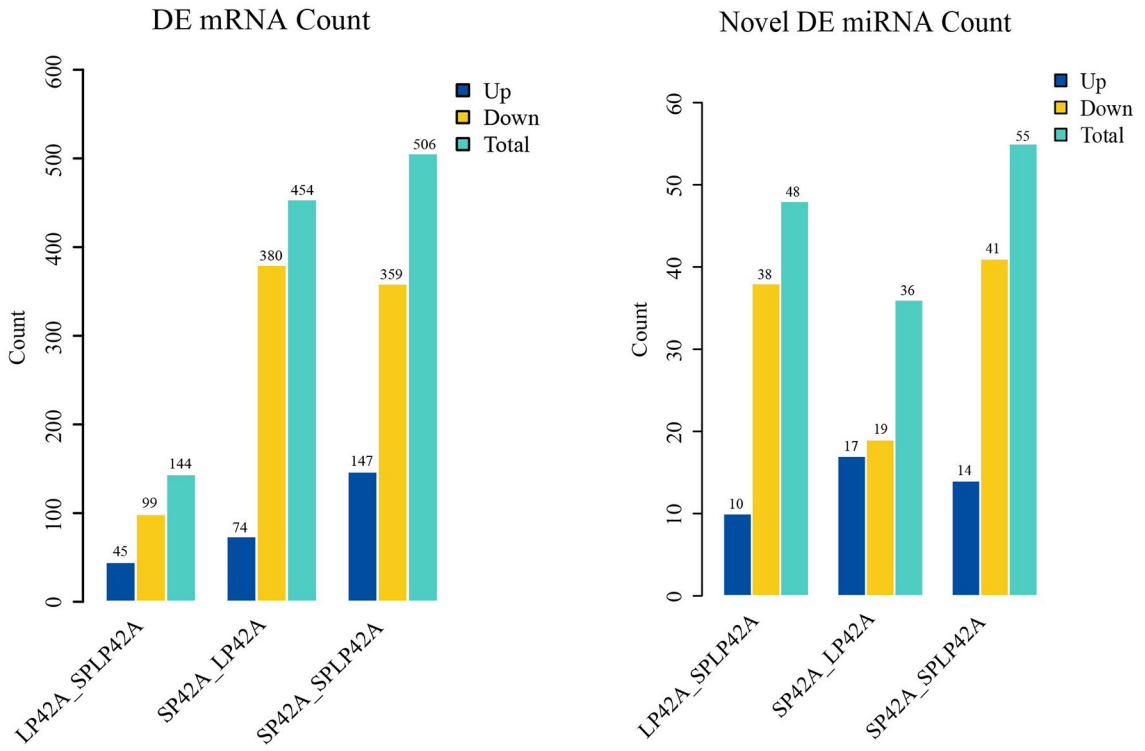


FIGURE 4
Differentially expressed (DEM) microRNA and (DEG) mRNA analysis. Different color rectangles were used to represent the numbers of upregulated (blue), downregulated (yellow) and total (green) differential genes.

precursor may give rise to functional levels of mature miRNA (35, 36). The dominant products may change from species to species and have different tissue expression preferences, including normal vs. pathological tissue (37–43). At the same time, when identifying three biological repetitive samples in each treatment group, we found that except for SP42-1 and SP42-3, all the other individuals identified 106 stem-loop sequences. The stem-loop sequences that were not detected were oar-mir-654 (producing 3P mature miRNA) and oar-mir-1193 (producing 5p mature miRNA). mir-1193 has been found to inhibit the proliferation and invasion of cancer cells by directly acting on transmembrane 9 superfamily member 3 (*TM9SF3*) and insulin-like growth factor 2 mRNA binding protein 2 (*IGF2BP2*) (44, 45). In a mouse study, overexpression of mir-1193-5p was found to be able to increase the differentiation tendency of oligodendrocyte progenitor cells (46). Similarly, previous studies have found that mir-654 is significantly related to the development of a variety of malignant tumors, including lung cancer, rectal cancer, esophageal cancer and colon cancer (47, 48). Since we only detected the above two miRNAs in SP42-2 individuals, it is not possible to determine their relationship with the photoperiod and adrenal gland, so we will explore further.

After differential expression analysis, we initially identified 144, 454 and 506 DEGs and 48, 36 and 55 DE miRNAs (LP42 vs. SPLP42, SP42 vs. LP42 and SP42 vs. SPLP42). According to the grouping, the quantity distribution of DEGs was more consistent with the expectation of the experiment; that is, the number of differentially expressed genes in the SP42 vs. LP42 and SP42 vs. SPLP42 comparison groups was greater than the number of differentially expressed genes in the LP42 vs. SPLP42 comparison group. However, the results of miRNA analysis showed that there were the fewest DEMs in the SP42 vs. LP42 comparison group. When studying the data of the sheep pituitary transcriptome after different photoperiod treatments, with the increase in long photoperiod maintenance days, the number of differentially expressed genes and differentially expressed lncRNAs were found to increase significantly compared with the short photoperiod (22). What causes the number of differential miRNAs to appear in the small probability of the significant difference treatment group needs further analysis. Similarly, during a miRNA study of sheep hypothalamus, significant expression of the let-7 and oar-miRNA-200 families was found (17), and some studies also proved that those identified miRNAs were differentially expressed in seasonal and non-seasonal sheep breeds (49).

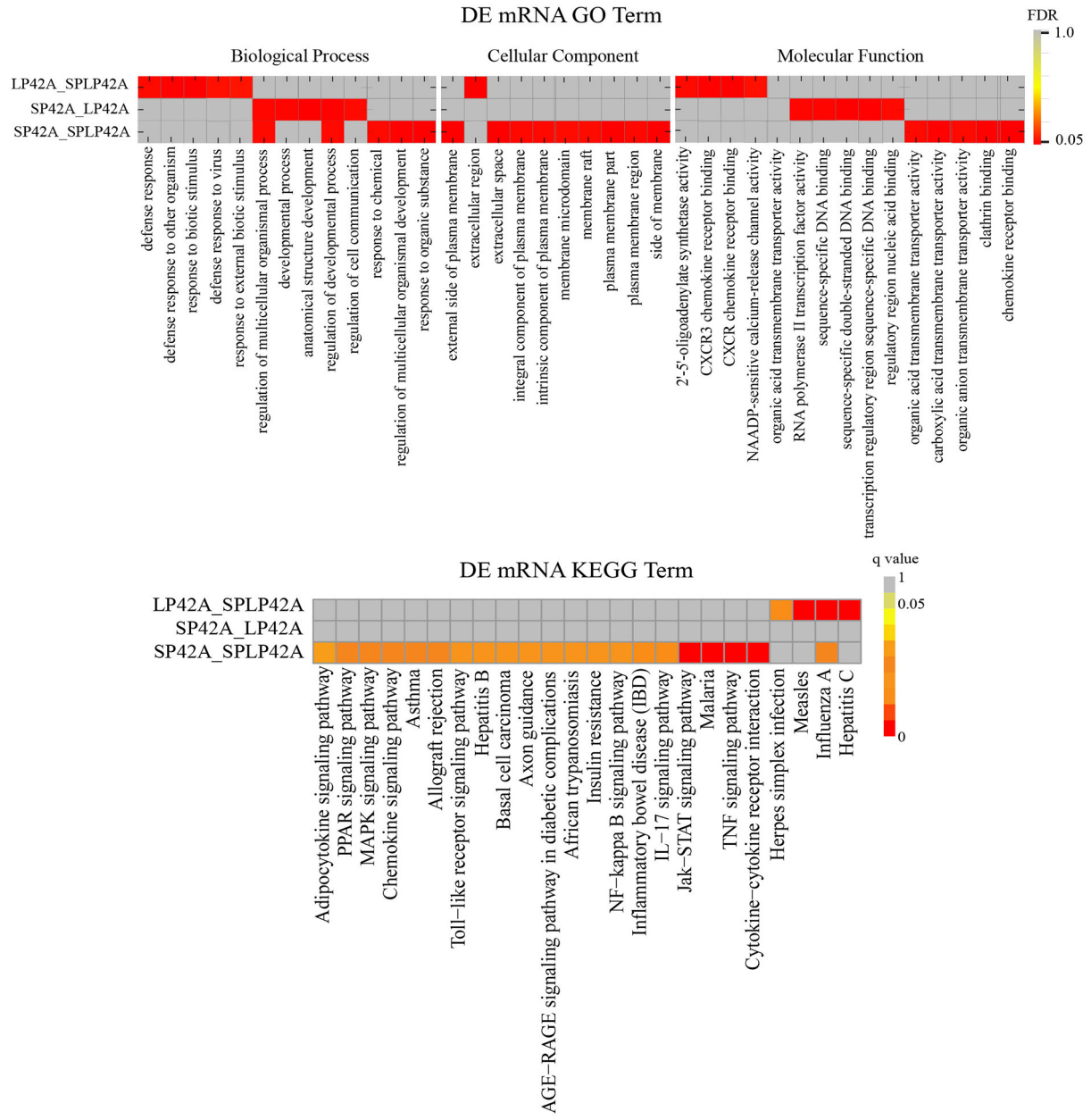


FIGURE 5
Functional enrichment analysis of DEGs. The upper part of the picture shows the enrichment results of GO and the lower part shows the enrichment results of KEGG.

However, it is unfortunate that although the let-7 and oar-miRNA-200 families were detected in this study, there was no differential expression in different comparison groups. In our known miRNA identification of sheep adrenal glands and Small Tail Han sheep hypothalamus (17), only six miRNAs (oar-miR-544-3p, oar-miR-411b-5p, oar-miR-376e-3p, oar-miR-376d, oar-miR-376b-3p, oar-miR-376a-3p) were specifically expressed in the adrenal gland, and one (oar-miR-1193-3p) was specifically expressed in the hypothalamus. The rest can be detected in both tissues. In addition, several miRNAs, such as miRNA-200 family members, were thought to be conserved in the hypothalamus of mice (16, 50), rats (51), zebrafish (52), and sheep (17). However, to sum up, it may be conserved in the species, not in sheep breeds and the tissue; perhaps it is just the wrong time to monitor.

All the DEGs were subjected to GO functional enrichment and KEGG enrichment analysis according to the set comparison group. Because there were too many enrichment items in the BP part and the MF part, we chose the top five terms that were

TABLE 3 Functional enrichment analysis of target gene of DEMs.

Target gene of know miRNA				Target gene of novel miRNA			
Term_name (LP42-SPLP42)	**Accession**	**FDR**	**Type**	**Term_name (SP42-SPLP42)**	**Accession**	**FDR**	**Type**
DNA methylation on cytosine within a CG sequence	GO:0010424	0.021	BP	oligo-1,6-glucosidase activity	GO:0004574	0.001	MF
Cerebellum vasculature morphogenesis	GO:0061301	0.021		alpha-1,4-glucosidase activity	GO:0004558	0.003	
Cellular response to lipid	GO:0071396	0.021		alpha-glucosidase activity	GO:0090599	0.006	
Progesterone secretion	GO:0042701	0.021		hydrolase activity, hydrolyzing O-glycosyl compounds	GO:0004553	0.009	
Steroid hormone secretion	GO:0035929	0.021		hydrolase activity, acting on glycosyl bonds	GO:0016798	0.015	
Cell surface	GO:0009986	0.047	CC	glucosidase activity	GO:0015926	0.015	
DNA-methyltransferase activity	GO:0009008	0.008	MF	lysozyme activity	GO:0003796	0.022	
DNA (cytosine-5-)-methyltransferase activity, acting on CpG substrates	GO:0051718	0.008		peptidoglycan muralytic activity	GO:0061783	0.033	
DNA (cytosine-5-)-methyltransferase activity	GO:0003886	0.008		rRNA (uridine-N3-)-methyltransferase activity	GO:0070042	0.039	
Unmethylated CpG binding	GO:0045322	0.009		beta-fructofuranosidase activity	GO:0004564	0.039	
Wnt-activated receptor activity	GO:0042813	0.035		UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase activity	GO:0003975	0.039	
Wnt-protein binding	GO:0017147	0.037		sucrose alpha-glucosidase activity	GO:0004575	0.039	
Cytokine binding	GO:0019955	0.044		phospho-N-acetylmuramoyl -pentapeptide-transferase activity	GO:0008963	0.039	

The above GO term includes the significantly enriched term in the analysis results of all the comparison groups, except for the BP group in the known miRNA section, which ranks in the top 5. The absence of the comparison group name indicates that the comparison group has no significant enrichment of term.

significantly enriched. In the BP section, we find that there are coincident entries GO:0051239 and GO:0050793 in the SP42 vs. LP42 and SP42 vs. SPLP42 comparison groups. The overlapping genes *GRHL2* (grainy head-like protein 2 homologs), *CENPF* (centromere protein F) and *FGF16* (fibroblast growth factor 16) were found by further analysis of the above two terms. *GRHL2* is one of three mouse homologs of *Drosophila Grainyhead* (53) and is expressed in diverse embryonic epithelial tissues during development (53, 54). *GRHL2* and its paralogue *Grhl3* play essential roles in neural tube closure in mice (55–58). Data indicate that a conserved *GRHL2*-coordinated gene network controls trophoblast branching morphogenesis, thereby facilitating the development of the site of feto-maternal exchange (59). Normal placental development is a key factor in reproductive success, so whether screening this gene in our different photoperiod comparison groups indicates that placental development is affected by photoperiod is unknown. Therefore, it is very important to explore its function in the process of seasonal reproduction of sheep. Another key gene, *CENPF*, related to early embryonic development was also screened out. *CENPF* is a member of the centromere protein family that regulates chromosome segregation during mitosis (60, 61). Zhou et al. used mice to study and indicated that farnesylation plays a key role during *CENPF* degradation and localization in early embryos. At the same time, a knockout test of the *CENPF* gene showed that the gene is very important for early embryos (62). The fibroblast growth factor 16 (*FGF16*) gene is correlated with oocyte quality (63). In summer, when oocyte quality is low, the expression of *FGF16* is low. Conversely, in winter, when oocyte quality is high, the expression of *FGF16* is high. We found that the expression of the *FGF16* gene was upregulated in the comparative group of SP42 vs. LP42, indicating that in adrenal tissue, the expression of the *FGF16* gene under a long photoperiod was nearly 10 times lower than that under short photoperiod.

Our experimental analysis results are not ideal, whether it is the distribution of differential miRNAs in different comparison groups or the functional enrichment analysis of GO and KEGG of DEGs or the target genes of DEMs. However, the key comparison group SP42 vs. LP42 we set up has only made exciting discoveries in the process of GO enrichment analysis of mRNA. All the significantly enriched KEGG items of DEGs

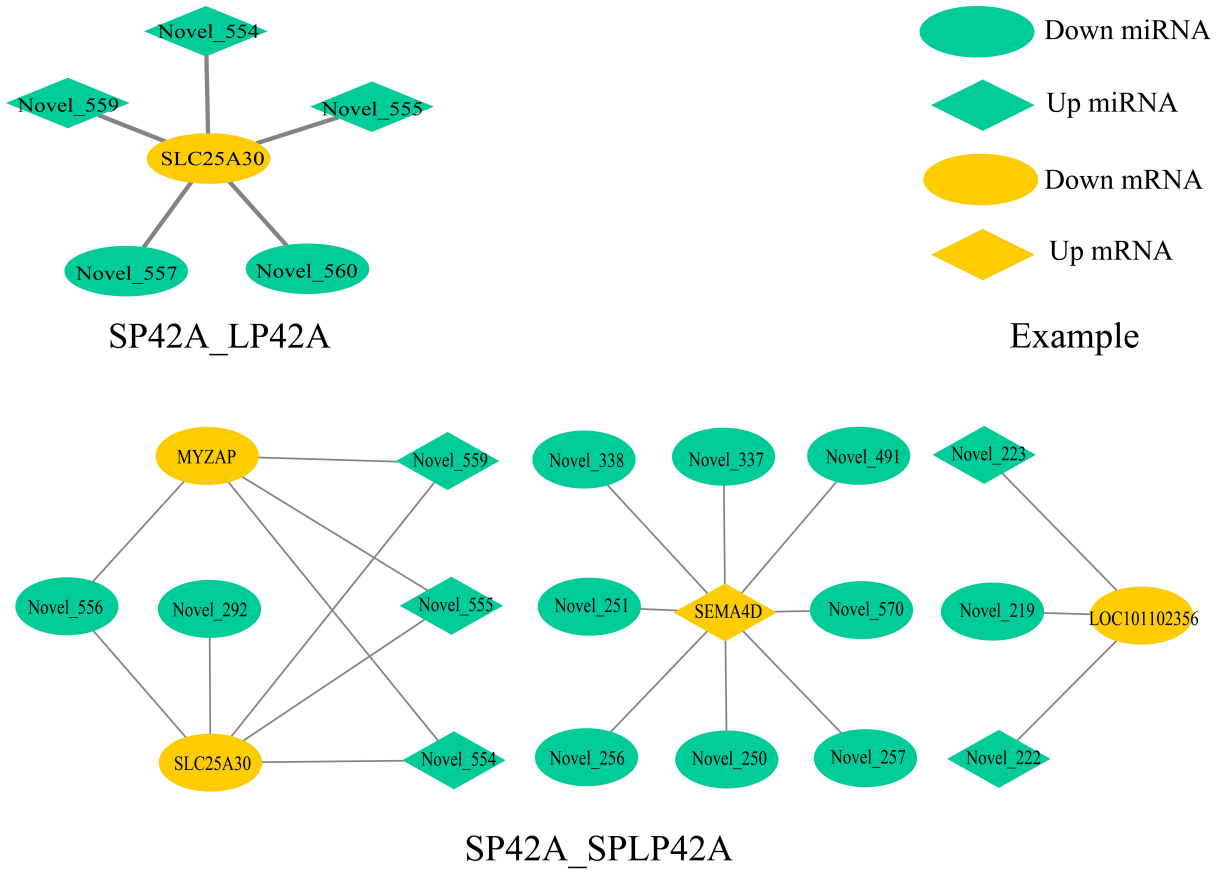


FIGURE 6
Overview of mRNA–miRNA networks. Different colors represent DEM (green) and DEG (yellow), and different shapes represent different modes of regulation.

or target genes of DEMs are shown in the results section, but the SP42 vs. LP42 comparison group unexpectedly did not have any significant entries. Therefore, we need to further construct a miRNA–mRNA coexpression network to study interaction regulation and mine gene expression regulation patterns. We identified the overlapping core gene *SLC25A30* in the SP42 vs. LP42 and SP42 vs. SPLP42 comparison groups. SLC25A30 is a member of the mitochondrial transporter family (64). Mitochondrial transporters for inorganic anions/malate (SLC25A30), thiamine pyrophosphate (TPP) (SLC25A19) and iron (SLC25A28) are also considered conceptus-induced IFNT-dependent endometrial DEGs and are increased in endometrial epithelial cells by IFNT (65, 66). A study by Gorgoglione et al. (67) suggests that SLC25A30 may have the function of exporting sulfite and thiosulfate or transporting malate acid, which may contribute to the flow of malate–aspartate acid in mitochondria. Importantly, malic acid and malate dehydrogenase is present in the intrauterine environment of cattle, and malic acid has been shown to affect the early embryonic development of hamsters (68–70). *SLC25A30* gene expression has been reported to be upregulated by cellular oxidative stress, and our study may also be a sign of increased oxidative stress (71). Novel miRNA554, novel miRNA555 and novel miRNA559 are the common miRNAs that regulate the core gene in the two comparison groups (SP42 vs. LP42 and SP42 vs. SPLP42). These three miRNAs are located on chromosome 7 and belong to the same miRNA cluster according to location analysis. However, the specific functions need to be further analyzed.

These results suggest that several key DEGs and DEMs in the adrenal gland are directly or indirectly involved in the process of photoperiod-changing reproductive activity, and further study of gene/miRNA knockout or overexpression is helpful for us to understand their real function in female reproductive traits.

Conclusions

We successfully obtained the mRNA and small RNA data of adrenal gland tissue samples of Sunite sheep under different photoperiod treatments by RNA sequencing

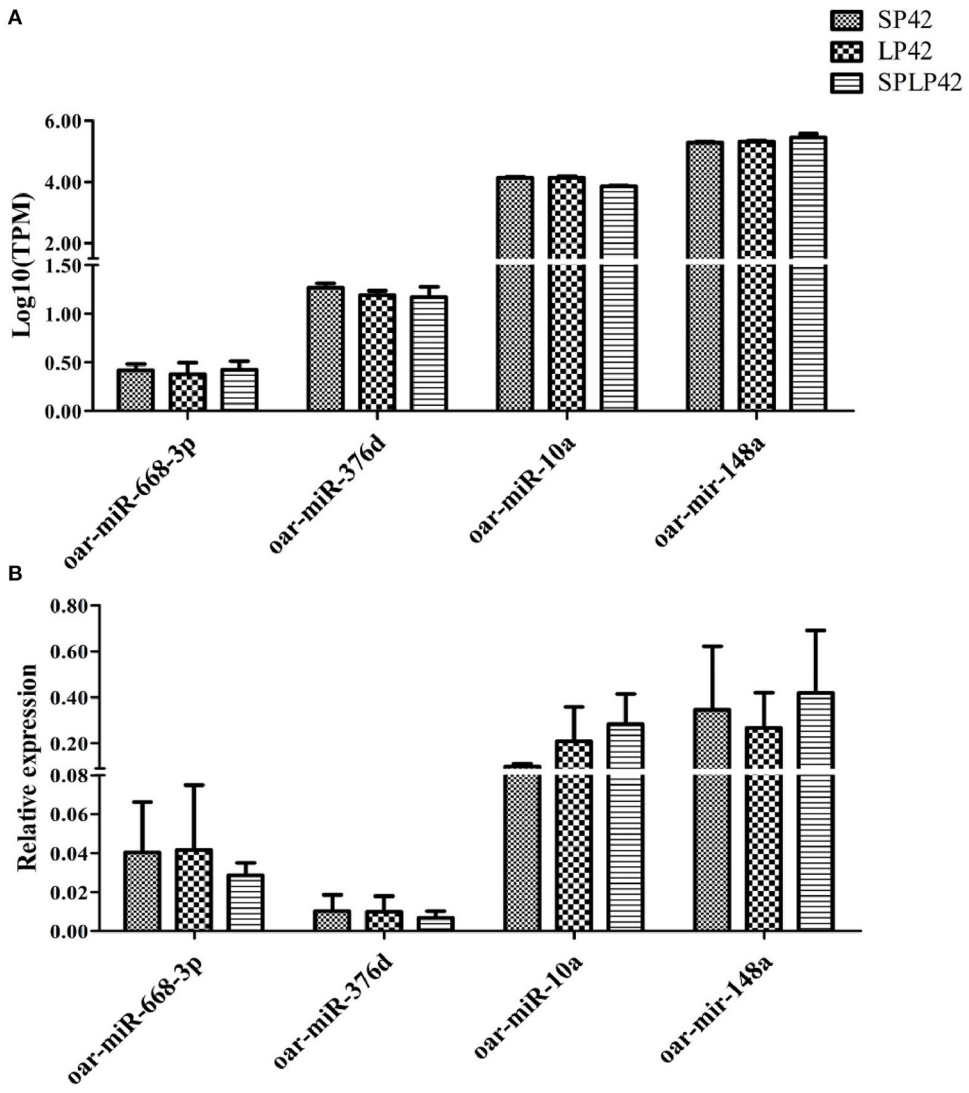


FIGURE 7
Quantitative verification results of miRNA. The relative expressions of 4 miRNA randomly selected in different individuals and among comparison groups were verified. Rectangles with different patterns represent RNA-seq **(A)** and qPCR **(B)** results, respectively.

and bioinformatics analysis. The key candidate genes of photoperiod affecting reproduction, such as *GRHL2*, *CENPF*, *FGF16* and *SLC25A30*, were confirmed. The miRNA (oar-miR-544-3p, oar-miR-411b-5p, oar-miR-376e-3p, oar-miR-376d, oar-miR-376b-3p, oar-miR-376a-3p) was specifically expressed in adrenal tissue. The predicted mRNA-miRNA pairs (*SLC25A30* regulated by Novel miRNA554, Novel miRNA555 and Novel miRNA559) showed significant differences in SP42 vs. the other two groups, indicating that the relationship may play an important role in the process of different photoperiod affecting adrenal function. Our results provide a new perspective for the study of sheep reproduction and help to deepen the understanding of ovine reproduction.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/sra/PRJNA756142; https://www.ncbi.nlm.nih.gov/sra/PRJNA811389.

Ethics statement

The animal study was reviewed and approved by the Animal Experimental Welfare Ethics Committee of the Institute of Animal Sciences, Chinese Academy of Agricultural Sciences.

Author contributions

QiuL and MC designed the research. XD wrote the paper and performed the study. RD, XD, and QinL collected the data. XD and XH analyzed data. MC revised the final manuscript. All authors reviewed the manuscript and approved the final version.

Funding

This work was financially supported by National Natural Science Foundation of China (32172704), China Agriculture Research System of MOF and MARA (CARS-38), and the Agricultural Science and Technology Innovation Program of China (CAAS-ZDRW202106 and ASTIP-IAS13).

Acknowledgments

We are grateful to Xiaosheng Zhang, Jinlong Zhang, and Xiaofei Guo in Tianjin Institute of Animal Sciences for the ovariectomy, estrogen implantation, and ewes feeding. Thanks to Xiaohan Cao, Xiaoyu Li, Xinlong Dong, and Qing Xia for their help in the sample collection.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.888207/full#supplementary-material

References

1. La YF, He XY, Zhang LP, Di R, Wang XY, Gan SQ, et al. Comprehensive analysis of differentially expressed profiles of mRNA, lncRNA, and circRNA in the uterus of seasonal reproduction sheep. *Genes-Basel.* (2020) 11:301–19. doi: 10.3390/genes11030301

2. Ebling FJP FD. Photoperiod requirements for puberty differ from those for the onset of the adult breeding season in female sheep. *J Reprod Fertil.* (1988) 84:283–93. doi: 10.1530/jrf.0.0840283

3. Dupre SM, Miedzinska K, Duval CV, Le Y, Goodman RL, Lincoln GA, et al. Identification of *Eya3* and *TAC1* as long-day signals in the sheep pituitary. *Curr Biol.* (2010) 20:829–35. doi: 10.1016/j.cub.2010.02.066

4. Li XY, He XY, Liu QY, Wang XY, Guo XF, Xia Q, et al. Expression pattern analysis of *TAC1* and *PRLR* genes in different reproductive states of sheep. *Acta Vet Zootech Sin.* (2018) 49:253–62. doi: 10.11843/j.issn.0366-6964.2018.02.004

5. Batailler M, Chesneau D, Derouet L, Butruille L, Segura S, Cognie J, et al. Pineal-dependent increase of hypothalamic neurogenesis contributes to the timing of seasonal reproduction in sheep. *Sci Rep.* (2018) 8:6188. doi: 10.1038/s41598-018-24381-4

6. Karsch FJ, Bittman EL, Foster DL, Goodman RL, Legan SJ, Robinson JE. Neuroendocrine basis of seasonal reproduction. *Recent Prog Horm Res.* (1984) 40:185–232. doi: 10.1016/B978-0-12-571140-1.50010-4

7. Weems PW, Goodman RL, Lehman MN. Neural mechanisms controlling seasonal reproduction: principles derived from the sheep model and its comparison with hamsters. *Front Neuroendocrinol.* (2015) 37:43–51. doi: 10.1016/j.yfrne.2014.12.002

8. Masumoto K, Ukai-Tadenuma M, Kasukawa T, Nagano M, Uno KD, Tsujino K, et al. Acute induction of *Eya3* by late-night light stimulation triggers *TSH* beta expression in photoperiodism. *Neurosci Res.* (2011) 71:E172. doi: 10.1016/j.neures.2011.07.742

9. Hanon EA, Lincoln GA, Fustin JM, Dardente H, Masson-Pevet M, Morgan PJ, et al. Ancestral *TSH* mechanism signals summer in a photoperiodic mammal. *Curr Biol.* (2008) 18:1147–52. doi: 10.1016/j.cub.2008.06.076

10. Acevedo-Rodriguez A, Kauffman AS, Cherrington BD, Borges CS, Roepke TA, Laconi M. Emerging insights into hypothalamic-pituitary-gonadal axis regulation and interaction with stress signalling. *J Neuroendocrinol.* (2018) 30:e12590. doi: 10.1111/jne.12590

11. Seckl JR, Holmes MC. Mechanisms of Disease: glucocorticoids, their placental metabolism and fetal'programming' of adult pathophysiology. *Nat Clin Pract Endoc.* (2007) 3:479–88. doi: 10.1038/ncpendmet0515

12. Pereira OCM, Arena AC, Yasuhara F, Kempinas WG. Effects of prenatal hydrocortisone acetate exposure on fertility and sexual behavior in male rats. *Regul Toxicol Pharm.* (2003) 38:36–42. doi: 10.1016/S0273-2300(03)00046-1

13. Piffer RC, Garcia PC, Pereira OCM. Adult partner preference and sexual behavior of male rats exposed prenatally to betamethasone. *Physiol Behav.* (2009) 98:163–7. doi: 10.1016/j.physbeh.2009.05.003

14. Hardy MP, Gao HB, Dong Q, Ge RS, Wang Q, Chai WR, et al. Stress hormone and male reproductive function. *Cell Tissue Res.* (2005) 322:147–53. doi: 10.1007/s00441-005-0006-2

15. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science.* (2014) 344:1168–73. doi: 10.1126/science.1252806

16. Correia CN, McLoughlin KE, Nalpas NC, Magee DA, Browne JA, Rue-Albrecht K, et al. RNA Sequencing (RNA-seq) reveals extremely low levels of reticulocyte-derived globin gene transcripts in peripheral blood from horses (*Equus caballus*) and cattle (*Bos taurus*). *Front Genet.* (2018) 9:278. doi: 10.3389/fgene.2018.00278

17. Zhang ZB, Tang JS, Di R, Liu QY, Wang XY, Gan SQ, et al. Integrated hypothalamic transcriptome profiling reveals the reproductive roles of mRNAs and miRNAs in Sheep. *Front Genet.* (2019) 10:1296. doi: 10.3389/fgene.2019.01296

18. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol.* (2019) 20:21–37. doi: 10.1038/s41580-018-0045-7

19. Lv XY, Chen WH, Wang SH, Cao XK, Yuan ZH, Getachew TF, et al. Integrated hair follicle profiles of microRNAs and mRNAs to reveal the pattern formation of Hu sheep lambskin. *Genes.* (2022) 13:342. doi: 10.3390/genes13020342

20. Chen S, Wang CQ, Chen QL, Zhao DT, Liu YB, Du L, et al. Downregulation of three novel miRNAs in the lymph nodes of sheep immunized with the brucella suis strain 2 vaccine. *Front Vet Sci.* (2022) 9:813170. doi: 10.3389/fvets.2022.813170

21. Fei XJ, Jin ML, Wang YQ, Li TT, Lu ZK, Yuan ZH, et al. Transcriptome reveals key microRNAs involved in fat deposition between different tail sheep breeds. *PLoS ONE.* (2022) 17:e0264804. doi: 10.1371/journal.pone.0264804

22. Xia Q, Chu MX, He XY, Zhang XS, Zhang JL, Guo XF, et al. Identification of photoperiod-induced lncRNAs and mRNAs in pituitary pars tuberalis of sheep. *Front Vet Sci.* (2021) 8:644474. doi: 10.3389/fvets.2021.753614

23. He XY, Tao L, Zhong YJ, Di R, Xia Q, Wang XY, et al. Chu MX: Photoperiod induced the pituitary differential regulation of lncRNAs and mRNAs related to reproduction in sheep. *PeerJ.* (2021) 9:e10953. doi: 10.7717/peerj.10953

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

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

26. Griffiths-Jones S. miRBase: the microRNA sequence database. *Methods Mol Biol.* (2006) 342:129–38. doi: 10.1385/1-59745-123-1:129

27. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* (2013) 41:D226. doi: 10.1093/nar/gks1005

28. Huda A, Jordan IK. Analysis of transposable element sequences using CENSOR and RepeatMasker. *Methods Mol Biol.* (2009) 537:323–36. doi: 10.1007/978-1-59745-251-9_16

29. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033

30. Friedlander MR, Mackowiak SD Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* (2012) 40:37–52. doi: 10.1093/nar/gkr688

31. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* (2010) 28:511–U174. doi: 10.1038/nbt.1621

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

33. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in *Drosophila*. *Genome Biol.* (2003) 5:R1. doi: 10.1186/gb-2003-5-1-r1

34. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

35. Du T, Zamore PD. MicroPrimer: the biogenesis and function of microRNA. *Development.* (2005) 132:4645–52. doi: 10.1242/dev.02070

36. Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol.* (2007) 8:23–36. doi: 10.1038/nrm2085

37. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. *EMBO Rep.* (2011) 12:172–7. doi: 10.1038/embor.2010.191

38. Ro S, Park C, Young D, Sanders KM, Yan W. Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res.* (2007) 35:5944–53. doi: 10.1093/nar/gkm641

39. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* (2007) 17:1850–64. doi: 10.1101/gr.6597907

40. De Wit E, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res.* (2009) 19:2064–74. doi: 10.1101/gr.093781.109

41. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* (2010) 24:992–1009. doi: 10.1101/gad.1884710

42. Marco A, Macpherson JI, Ronshaugen M, Griffiths-Jones S. MicroRNAs from the same precursor have different targeting properties. *Silence.* (2012) 3:8. doi: 10.1186/1758-907X-3-8

43. Li SC, Tsai KW, Pan HW, Jeng YM, Ho MR, Li WH. MicroRNA 3' end nucleotide modification patterns and arm selection preference in liver tissues. *BMC Syst Biol.* (2012) 6:S14. doi: 10.1186/1752-0509-6-S2-S14

44. Li XL, Li YH, Lu H. MiR-1193 suppresses proliferation and invasion of human breast cancer cells through directly targeting *IGF2BP2*. *Oncol Res.* (2017) 25:579–85. doi: 10.3727/97818823455816X14760504645779

45. Shen LY, Du XJ, Ma HY, Mei SX. MiR-1193 suppresses the oroliferation and invasion of human T-cell leukemia cells through directly targeting the transmembrane 9 superfamily 3 (*TM9SF3*). *Oncol Res.* (2017) 25:1643–51. doi: 10.3727/096504017X14908284471361

46. Ma Q, Matsunaga A, Ho B, Oksenberg JR, Didonna A. Oligodendrocyte-specific Argonaute profiling identifies microRNAs associated with experimental autoimmune encephalomyelitis. *J Neuroinflamm.* (2020) 17:297. doi: 10.1186/s12974-020-01964-5

47. Pu JT, Hu Z, Zhang DG, Zhang T, He KM, Dai TY. MiR-654-3p suppresses non-small cell lung cancer tumourigenesis by inhibiting *PLK4*. *Oncotargets Ther.* (2020) 13:7997–8008. doi: 10.2147/OTT.S258616

48. Li P, Cai JX, Han F, Wang J, Zhou JJ, Shen KW, et al. Expression and significance of miR-654-5p and miR-376b-3p in patients with colon cancer. *World J Gastro Oncol.* (2020) 12:492–502. doi: 10.4251/wjgo.v12.i4.492

49. Zhai M, Xie Y, Liang H, Lei X, Zhao Z. Comparative profiling of differentially expressed microRNAs in estrous ovaries of Kazakh sheep in different seasons. *Gene.* (2018) 664:181–91. doi: 10.1016/j.gene.2018.04.025

50. Choi PS, Zakhary L, Choi WY, Caron S, Alvarez-Saavedra E, Miska EA, et al. Members of the miRNA-200 family regulate olfactory neurogenesis. *Neuron.* (2008) 57:41–55. doi: 10.1016/j.neuron.2007.11.018

51. Sangiao-Alvarellos S, Pena-Bello L, Manfredi-Lozano M, Tena-Sempere M, Cordido F. Perturbation of hypothalamic microRNA expression patterns in male rats after metabolic distress: impact of obesity and conditions of negative energy balance. *Endocrinology.* (2014) 155:1838–50. doi: 10.1210/en.2013-1770

52. Garaffo G, Conte D, Provero P, Tomaiuolo D, Luo Z, Pinciroli P, et al. The *Dlx5* and *Foxg1* transcription factors, linked via miRNA-9 and−200, are required for the development of the olfactory and GnRH system. *Mol Cell Neurosci.* (2015) 68:103–19. doi: 10.1016/j.mcn.2015.04.007

53. Wilanowski T, Tuckfield A, Cerruti L, O'Connell S, Saint R, Parekh V, et al. A highly conserved novel family of mammalian developmental transcription factors related to *Drosophila* grainyhead. *Mech Dev.* (2002) 114:37–50. doi: 10.1016/S0925-4773(02)00046-1

54. Auden A, Caddy J, Wilanowski T, Ting SB, Cunningham JM, Jane SM. Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. *Gene Expr Patterns.* (2006) 6:964–70. doi: 10.1016/j.modgep.2006.03.011

55. Rifat Y, Parekh V, Wilanowski T, Hislop NR, Auden A, Ting SB, et al. Regional neural tube closure defined by the Grainy head-like transcription factors. *Dev Biol.* (2010) 345:237–45. doi: 10.1016/j.ydbio.2010.07.017

56. Pyrgaki C, Liu A, Niswander L. Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Dev Biol.* (2011) 353:38–49. doi: 10.1016/j.ydbio.2011.02.027

57. Brouns MR, De Castro SC, Terwindt-Rouwenhorst EA, Massa V, Hekking JW, Hirst CS, et al. Over-expression of *GRHL2* causes spina bifida in the Axial defects mutant mouse. *Hum Mol Genet.* (2011) 20:1536–46. doi: 10.1093/hmg/ddr031

58. Werth M, Walentin K, Aue A, Schonheit J, Wuebken A, Pode-Shakked N, et al. The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development.* (2010) 137:3835–45. doi: 10.1242/dev.055483

59. Walentin K, Hinze C, Werth M, Haase N, Varma S, Morell R, et al. A *GRHL2*-dependent gene network controls trophoblast branching morphogenesis. *Development.* (2015) 142:1125–36. doi: 10.1242/dev.113829

60. Wan X, O'Quinn RP, Pierce HL, Joglekar AP, Gall WE, DeLuca JG, et al. Protein architecture of the human kinetochore microtubule attachment site. *Cell.* (2009) 137:672–84. doi: 10.1016/j.cell.2009.03.035

61. Karpen GH, Allshire RC. The case for epigenetic effects on centromere identity and function. *Trends Genet.* (1997) 13:489–96. doi: 10.1016/S0168-9525(97)01298-5

62. Zhou CJ, Wang XY, Han Z, Wang DH, Ma YZ, Liang CG. Loss of *CENPF* leads to developmental failure in mouse embryos. *Cell Cycle.* (2019) 18:2784–99. doi: 10.1080/15384101.2019.1661173

63. Ferreira RM, Chiaratti MR, Macabelli CH, Rodrigues CA, Ferraz ML, Watanabe YF, et al. The infertility of repeat-breeder cows during summer is associated with decreased mitochondrial DNA and increased expression of mitochondrial and apoptotic genes in oocytes. *Biol Reprod.* (2016) 94:1–10. doi: 10.1095/biolreprod.115.133017

64. Hediger MA, Clemencon B, Burrier RE, Bruford EA. The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Mol Aspects Med.* (2013) 34:95–107. doi: 10.1016/j.mam.2012.12.009

65. Mathew DJ, Sanchez JM, Passaro C, Charpigny G, Behura SK, Spencer TE, et al. Interferon tau-dependent and independent effects of the bovine

conceptus on the endometrial transcriptome. *Biol Reprod.* (2019) 100:365–80. doi: 10.1093/biolre/ioy199

66. Ruprecht JJ, Kunji ERS. The SLC25 mitochondrial carrier family: structure and mechanism. *Trends Biochem Sci.* (2020) 45:244–58. doi: 10.1016/j.tibs.2019.11.001

67. Gorgoglione R, Porcelli V, Santoro A, Daddabbo L, Vozza A, Monne M, et al. The human uncoupling proteins 5 and 6 (UCP5/SLC25A14 and UCP6/SLC25A30) transport sulfur oxyanions, phosphate and dicarboxylates. *Biochim Biophys Acta Bioenerg.* (2019) 1860:724–33. doi: 10.1016/j.bbabio.2019.07.010

68. Forde N, Simintiras CA, Sturmey R, Mamo S, Kelly AK, Spencer TE, et al. Amino acids in the uterine luminal fluid reflects the temporal changes in transporter expression in the endometrium and conceptus during early pregnancy in cattle. *PLoS ONE.* (2014) 9:e100010. doi: 10.1371/journal.pone.0100010

69. Tribulo P, Balzano-Nogueira L, Conesa A, Siqueira LG, Hansen PJ. Changes in the uterine metabolome of the cow during the first 7 days after estrus. *Mol Reprod Dev.* (2019) 86:75–87. doi: 10.1002/mrd.23082

70. Chaney HL, Grose LF, Charpigny G, Behura SK, Sheldon IM, Cronin JG, et al. Conceptus-induced, interferon tau-dependent gene expression in bovine endometrial epithelial and stromal cells. *Biol Reprod.* (2021) 104:669–83. doi: 10.1093/biolre/ioaa226

71. Haguenauer A, Raimbault S, Masscheleyn S, Gonzalez-Barroso MD, Criscuolo F, Plamondon J, et al. A new renal mitochondrial carrier, *KMCP1*, is up-regulated during tubular cell regeneration and induction of antioxidant enzymes. *J Biol Chem.* (2005) 280:22036–43. doi: 10.1074/jbc.M412136200

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/fgene.2022.923766

OPEN ACCESS

EDITED BY
Yang Zhou,
Huazhong Agricultural University, China

REVIEWED BY
Rostam Abdollahi-Arpanahi,
University of Georgia, United States
George R Wiggans,
Council on Dairy Cattle Breeding,
United States

*CORRESPONDENCE
Christine F. Baes,
cbaes@uoguelph.ca

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

RECEIVED 19 April 2022
ACCEPTED 14 July 2022
PUBLISHED 24 August 2022

CITATION
Makanjuola BO, Abdalla EA, Wood BJ and Baes CF (2022), Applicability of single-step genomic evaluation with a random regression model for reproductive traits in turkeys (Meleagris gallopavo).
Front. Genet. 13:923766.
doi: 10.3389/fgene.2022.923766

Applicability of single-step genomic evaluation with a random regression model for reproductive traits in turkeys (*Meleagris gallopavo*)

Bayode O. Makanjuola[1], Emhimad A. Abdalla[1], Benjamin J. Wood[1,2,3] and Christine F. Baes[1,4]*

[1]Centre for Genomic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, [2]School of Veterinary Science, University of Queensland, Gatton, QLD, Australia, [3]Hybrid Turkeys, Kitchener, ON, Canada, [4]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Fertility and hatchability are economically important traits due to their effect on poult output coming from the turkey hatchery. Traditionally, fertility is recorded as the number of fertile eggs set in the incubator (FERT), defined at a time point during incubation by the identification of a developing embryo. Hatchability is recorded as either the number of fertile eggs that hatched (hatch of fertile, HOF) or the number hatched from all the eggs set (hatch of set, HOS). These traits are collected throughout the productive life of the bird and are conventionally cumulated, resulting in each bird having a single record per trait. Genetic evaluations of these traits have been estimated using pedigree relationships. However, the longitudinal nature of the traits and the availability of genomic information have renewed interest in using random regression (RR) to capture the differences in repeatedly recorded traits, as well as in the incorporation of genomic relationships. Therefore, the objectives of this study were: 1) to compare the applicability of a RR model with a cumulative model (CUM) using both pedigree and genomic information for genetic evaluation of FERT, HOF, and HOS and 2) to estimate and compare predictability from the models. For this study, a total of 63,935 biweekly FERT, HOF, and HOS records from 7,211 hens mated to 1,524 toms were available for a maternal turkey line. In total, 4,832 animals had genotypic records, and pedigree information on 11,191 animals was available. Estimated heritability from the CUM model using pedigree information was 0.11 ± 0.02, 0.24 ± 0.02, and 0.24 ± 0.02 for FERT, HOF, and HOS, respectively. With random regression using pedigree relationships, heritability estimates were in the range of 0.04–0.09, 0.11–0.17, and 0.09–0.18 for FERT, HOF, and HOS, respectively. The incorporation of genomic information increased the heritability by an average of 28 and 23% for CUM and RR models, respectively. In addition, the incorporation of genomic information caused predictability to increase by approximately 11 and 7% for HOF and HOS, respectively; however, a decrease in predictability of about 12% was observed for FERT. Our findings suggest that RR models using pedigree and genomic relationships simultaneously will achieve a higher predictability than the traditional CUM model.

KEYWORDS
pedigree and genomics, random regression analysis, hatchability, fertility, turkeys

Introduction

Turkey meat continues to be a popular meat for consumption with a total production of approximately 6 million tonnes in 2019 (FAO, 2021). The continuous production of turkey poults is dependent on reproductive efficiency. Therefore, improvement of reproductive efficiency increases the production of turkey poults and has a direct effect on the economic growth of the industry. Given the importance of reproductive efficiency in turkeys, research emphasis has focused more on improving egg production (Nestor et al., 1996; Kranis et al., 2007), with less emphasis placed on fertility and hatchability traits. In contrast to egg production, fertility and hatchability traits in turkeys are more directly related to the production of poults. Furthermore, these traits are significant in improving reproductive efficiency in turkeys as they are easily and regularly collected over the productive life of the bird, as well as being influenced by genetic and environmental factors (Wolc and Olori, 2009).

Traditionally, genetic evaluation of these traits has been performed using cumulative (CUM) records collected over the productive life of the bird (Case et al., 2010). However, the longitudinal nature of these traits allows the opportunity to use a model that accounts for the differences in records collected at different time points. For longitudinal traits, random regression (RR) has often been proposed as the model of choice to better account for genetic and environmental variances at different time points. The first potential and practical application of RR was implemented on test-day milk yield in dairy cattle (Schaeffer and Dekkers, 1994). Since its first application, RR has been used to estimate genetic parameters for carcass conformation in beef cattle (Englishby et al., 2016), egg production and body weight in chicken (Anang et al., 2002; Rovadoscki et al., 2016), survival rate in dairy cattle (Sasaki et al., 2015), and body weight in goats (Kheirabadi and Rashidi, 2016). Furthermore, a broiler breeder study by Makanjuola et al. (2021) reported higher genetic gain with the RR model than the cumulative model for hatch of fertile trait.

The availability of genomic information in many species has allowed for a better estimation of the relationships between individuals. The combination of this information with pedigree information simultaneously in a single-step genomic best linear unbiased prediction (ssGBLUP) (Legarra et al., 2009; Misztal et al., 2009) has been shown to outperform the traditional pedigree BLUP approach (Aguilar et al., 2010; Christensen and Lund, 2010). Based on the ssGBLUP approach, Abdalla et al. (2019) observed a 16% increased accuracy for walking score in turkeys over traditional pedigree BLUP. In addition, Emamgholi Begli et al. (2021) observed that accuracies obtained with RR ssGBLUP were generally equal to or higher than those obtained with RR-PBLUP for egg production traits. Similarly, Oliveira et al. (2019) reported higher validation reliabilities for genomic estimated breeding values (GEBV) in comparison to parent averages for milk production traits in dairy cattle when using a multi-trait RR test-day model. Given the benefits of increasing the prediction accuracy with genomic information, as well as the limited number of RR ssGBLUP studies in turkeys, the aims of this study were to 1) estimate genetic parameters for FERT, HOS, and HOF in turkeys using CUM and RR models with pedigree and genomic information and 2) compare the predictive ability of CUM and RR models when using pedigree and genomic information.

Materials and methods

Phenotypes and pedigree data

Phenotypic data for this study were provided by Hybrid Turkeys, Kitchener, Canada. In total, 63,935 egg production records collected on a biweekly basis from a purebred turkey female line were available from 2010 to 2019 (Tables 1, 2). These egg production records were collected from 7,211 hens such that there was a total of 7,211 cumulative records, which indicates one record per hen. Cumulative record for each hen was calculated as the total number of eggs produced throughout the productive life of the hen, and this was subsequently used to derive the cumulative fertility and hatchability records. The 7,211 hens were mated to 1,524 toms, and eggs were set in the incubator biweekly throughout the productive life of the hen between 38 and 62 weeks of age. Individual hens were artificially inseminated, and trap-nest collected eggs were labeled with the identity of the hen. Consequently, this provided a pedigree for the progeny, as well as identification for the following hatchery traits. Fertility (FERT) records were collected by a process called candling, whereby light is passed through the eggs to determine the presence of a developing embryo. Hence, FERT was measured as the percent proportion of fertile eggs over the total egg set. Following the collection of fertility records, records of a successful hatching of an egg were used to calculate hatch of set (HOS), which is the proportion of all egg sets that hatched. Finally hatch of fertile (HOF) was defined as the proportion of fertile eggs that successfully hatched. Pedigree data for all animals with phenotypic records were provided and consisted of 15 generations and 11,191 individuals.

Genotype data

Of the animals with phenotypic records, a total of 4,832 animals were genotyped using a proprietary 65K SNP

TABLE 1 Descriptive statistics of the evaluated traits hatch of set (HOS), hatch of fertile (HOF), and fertility of set (FERT) including the number of records for each model, mean, standard deviation, and the number of records in the training and validation populations for each model.

Model	Trait	Mean ± SD	Number of records	Training population	Validation population
Cumulative (CUM)	HOS	67.94 ± 16.11			
	HOF	80.86 ± 14.52	7,211	6,447	764
	FERT	81.27 ± 14.02			
Random regression (RR)	HOS	68.64 ± 27.20			
	HOF	81.10 ± 26.07	63,935	56,471	7,464
	FERT	82.03 ± 22.75			

TABLE 2 Descriptive statistics for hatch of set (HOS), hatch of fertile (HOF), and fertility of set (FERT) including the number of records, mean, standard deviation (SD), and coefficient of variation (CV) for the different time points (38–62 weeks).

Week	Number of records	FERT ± SD	(FERT) CV%	HOF ± SD	(HOF) CV%	HOS ± SD	(HOS) CV%
38	2,838	84.06 ± 21.06	25.06	83.57 ± 25.49	30.50	71.53 ± 27.20	38.03
40	5,863	83.21 ± 19.47	23.39	83.69 ± 22.25	26.58	70.72 ± 24.70	34.92
42	6,395	82.29 ± 19.07	23.18	83.79 ± 21.02	25.08	69.97 ± 23.44	33.50
44	5,817	83.70 ± 19.06	22.78	83.80 ± 21.68	25.87	71.20 ± 23.84	33.49
46	5,687	84.06 ± 19.66	23.38	82.57 ± 23.21	28.11	70.76 ± 24.71	34.92
48	5,649	84.10 ± 20.55	24.44	82.42 ± 23.90	28.99	71.02 ± 25.36	35.71
50	5,507	83.84 ± 21.89	26.11	81.87 ± 24.99	30.53	70.56 ± 26.48	37.53
52	5,571	81.97 ± 23.84	29.08	81.40 ± 26.28	32.28	69.04 ± 27.74	40.18
54	4,781	81.83 ± 24.10	29.45	80.16 ± 27.37	34.14	67.85 ± 28.42	41.88
56	4,544	81.19 ± 24.76	30.50	79.82 ± 28.00	35.08	67.48 ± 28.85	42.75
58	4,348	79.96 ± 25.71	32.20	77.00 ± 30.25	39.28	64.73 ± 30.11	46.51
60	3,709	77.92 ± 27.61	35.43	75.82 ± 31.56	41.62	62.63 ± 31.16	49.75
62	3,226	74.48 ± 30.02	40.31	73.55 ± 34.40	46.77	59.33 ± 32.62	54.99

panel (Illumina, Inc). The genotype call rate was 94%, and missing genotypes were imputed using AlphaImpute version 2 (Whalen and Hickey, 2020). For imputation, both pedigree and population algorithms were used with a reference population of 1,626 animals. Default settings in AlphaImpute were used for imputation; however, the peeling and phasing cycles were increased to 50 cycles for pedigree and population algorithms, respectively. Increasing the phasing and peeling cycles was performed as a measure to increase the probability of achieving high-confidence-phased haplotypes and correctly calling the genotypes. An error rate of 0.01% was allowed for genotype calls. Imputation accuracy was estimated using allelic r^2, which is less dependent on allele frequencies and was greater than 98%. For quality control measures, non-autosomal SNP markers and autosomal SNP markers with MAF less than 0.05 and significantly deviating from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-8}$) were excluded. After editing, there were a total of 35,751 SNP markers retained for further analysis.

Statistical analysis

Best linear unbiased prediction

To investigate the influence of different parameters on reproductive traits, a CUM animal mixed model using only a single record per animal and a RR animal mixed model using all available records per animal were applied to estimate genetic parameters based only on pedigree relationships.

Cumulative model

With the CUM model, the following mixed model equation is used to estimate genetic parameters for reproductive traits:

$$\boldsymbol{y_{ij}} = \mu + \boldsymbol{hw_i} + \boldsymbol{a_j} + \boldsymbol{e_{ij}},$$

where $\boldsymbol{y_{ij}}$ is a vector of the CUM record of either FERT or HOF or HOS for the jth animal belonging to the ith hatch week, μ is the overall mean, $\boldsymbol{hw_i}$ is a vector of the fixed effect of the ith hatch week, and $\boldsymbol{a_j}$ is a vector of the random genetic effect of the animal. The assumption of the random effects was: $a_j \sim N(0, \boldsymbol{A}\sigma^2_{a_j})$,

where $\sigma^2_{a_j}$ is the additive genetic variance of the animal, $\sigma^2_{e_{ij}}$ is the error variance, and $\boldsymbol{A}$ is the numerator relationship matrix.

Random regression model

For the RR model, the following mixed model equation is used to estimate genetic parameters for reproductive trait biweekly records:

$$\boldsymbol{y_{ijkno}} = \mu + \boldsymbol{hw_i} + \boldsymbol{ehw_j} + \sum_{l=1}^{3} \boldsymbol{b_l X_{lk}} + \sum_{l=0}^{3} \boldsymbol{a_{Fnol} Z_{lk}} + \sum_{l=0}^{3} \boldsymbol{pe_{Fnol} Z_{lk}} + \sum_{l=0}^{1} \boldsymbol{a_{Mol} Z_{lk}} + \sum_{l=0}^{2} \boldsymbol{pe_{Mol} Z_{lk}} + \boldsymbol{e_{ijkno}},$$

where $\boldsymbol{y_{ijkno}}$ is a vector of repeated biweekly records for either FERT or HOF or HOS of the nth hen mated to the oth tom at the kth age belonging to the ith hatch week, μ is the overall mean, $\boldsymbol{hw_i}$ is a vector of the fixed effect of the ith hatch week, $\boldsymbol{ehw_j}$ is a vector of the fixed effect of the jth egg hatch week, $\boldsymbol{b_l}$ is a fixed regression coefficient of age of the hen when records were collected, $\boldsymbol{X_{lk}}$ is the incidence matrix value of the lth degree Legendre polynomial fitted for the effect of kth age, $\boldsymbol{a_{Fnol}}$ is the RR coefficient of additive genetic effect of the nth hen mated to the oth tom, $\boldsymbol{a_{Mol}}$ is the RR of additive genetic effect of the oth tom, $\boldsymbol{pe_{Fnol}}$ is the RR coefficient of permanent environment effect of the hen mated to the oth tom, $\boldsymbol{pe_{Mol}}$ is the RR coefficient of permanent environment effect of the oth tom, $\boldsymbol{Z_{lk}}$ is the incidence matrix value of the lth degree Legendre polynomial fitted for the additive genetic and permanent environment effects at kth age, and $\boldsymbol{e_{ijkno}}$ is the residual error term. The assumptions of the random effects were: $a_{Fno} \sim N(0, \boldsymbol{A}\sigma^2_{a_{Fno}})$, $a_{Mo} \sim N(0, \boldsymbol{A}\sigma^2_{a_{Mo}})$, $pe_{Mo} \sim N(0, \boldsymbol{I}\sigma^2_{pe_{Mo}})$, $pe_{Fno} \sim N(0, \boldsymbol{I}\sigma^2_{pe_{Fno}})$, and $e_{ijkno} \sim N(0, \boldsymbol{I}\sigma^2_{e_{ijkno}})$, where $\sigma^2_{a_{Fno}}$ is the hen additive genetic variance, $\sigma^2_{a_{Mo}}$ is the tom additive genetic variance, $\sigma^2_{pe_{Mo}}$ is the tom permanent environment variance, $\sigma^2_{pe_{Fno}}$ is the hen permanent environment variance, $\sigma^2_{e_{ijkno}}$ is the error variance, and $\boldsymbol{A}$ is the numerator relationship matrix. Also, the model was fitted using heterogeneous residuals per age class. For the purpose of comparison between CUM and RR models, variance components estimated for biweekly ages with the RR model were averaged to produce a single value for the additive genetic variance, permanent environment variance, and repeatability.

Single-step genomic best linear unbiased prediction

Following the presentation of creating a relationship matrix that included pedigree and genomic information (Legarra et al., 2009), an H relationship matrix derived from the combination of pedigree and genomic data was created to replace the pedigree relationship matrix (A) used in the aforementioned animal mixed model. Due to the computational cost of computing the H matrix, the inverse of H matrix is computed with a simpler structure:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (0.95\mathbf{G} + 0.05\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix, $\mathbf{A}_{22}^{-1}$ is the inverse of the $\mathbf{A}$ matrix of only the genotyped animals, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix estimated using the method presented by VanRaden (2008). Singular matrices are not invertible; therefore, to ensure that the G matrix is invertible, 0.05 of $\mathbf{A}_{22}$ was added to 0.95 of G. These weighting parameters were chosen because Abdalla et al. (2019) observed slightly more improvement in breast meat yield in turkeys with these weightings.

Estimates of the RR coefficients from BLUP and ssGBLUP RR models were used to derive the pedigree-estimated breeding value (EBV) and genomic-estimated breeding value (GEBV), respectively.

$$EBV_i = \boldsymbol{W}\hat{\alpha}_i,$$

$$GEBV_i = \boldsymbol{W}\hat{\delta}_i,$$

where $\hat{\alpha}_i$ is the estimated additive genetic regression coefficients for the ith animal, $\hat{\delta}_i$ is the estimated additive genomic regression coefficients for the ith animal, and $\boldsymbol{W}$ is a matrix of age covariate ranging from 38 to 62 weeks and associated with the degree of Legendre polynomials used.

TABLE 3 Summary of the fitted degree of Legendre polynomials[a] used in the random regression model with their corresponding log likelihoods (logL), Akaike information criteria (AIC), and the significance of their log likelihood ratio test (LRT) relative to the full model.

Model	a_F	a_M	pe_F	pe_M	logL	AIC	LRT
1	3	2	3	2	-231144.99	462382	Full model
2[b]	3	1	3	2	-231145.36	462376.71	NS
3	3	1	3	1	-231183.22	462446.43	[c]

a_F = additive genetic effect of the hen; a_M = additive genetic effect of the tom; pe_F = permanent environment effect of the hen; pe_M = permanent environment effect of the tom.
[a]Degree of Legendre polynomials: 1 = linear regression; 2 = quadratic regression; 3 = cubic regression.
[b]Model with the best fit in comparison to the full model.
[c]p value <0.001; NS: no significant difference; full model: model that had the highest number of parameters indicated by higher degree of Legendre polynomials.

Variance components and model comparison

All variance components and genetic parameters used in this study were estimated using the WOMBAT software program (Meyer, 2007). To adequately capture the parameters that

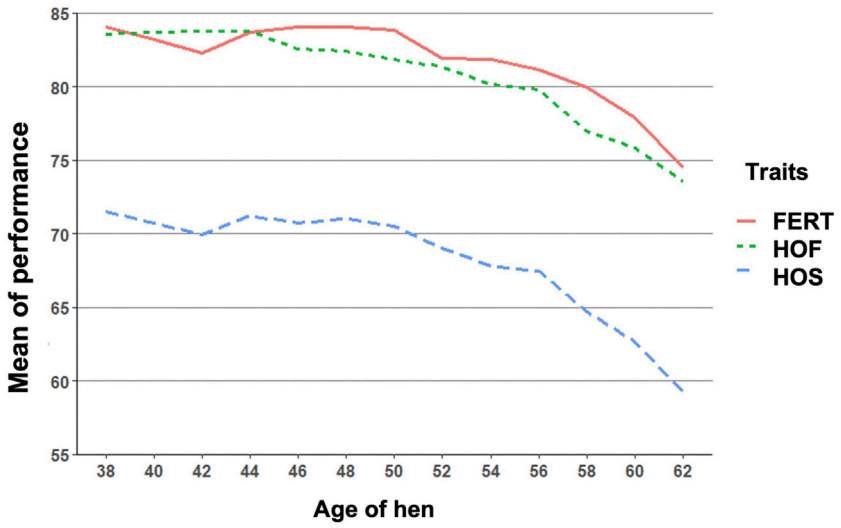


FIGURE 1
Biweekly mean of fertility of set (FERT), hatch of fertile (HOF), and hatch of set (HOS) from 38 to 62 weeks of hen age when records were measured.

contribute to the variation observed in reproductive traits when using RR models, three different models with varying degrees of Legendre polynomials were compared as shown in Table 3. The first model was a full model that had the highest possible degree of Legendre polynomials that converged. The second and third models were reduced models with lower degrees of Legendre polynomials. The criteria used to choose the best model were based on the log likelihood ratio test, the most parsimonious model that converged with both pedigree and genomic information and the Akaike information criterion (Akaike, 1998). For the fixed effect, a cubic polynomial was used because it appropriately described the trend in the biweekly reproductive traits as shown in Figure 1.

Predictive ability

Predictive ability of BLUP and ssGBLUP for all models was assessed using the following technique. Initially, observed phenotypes were corrected for all fixed effects fitted in the full model based on traditional BLUP (adjusted phenotype). Next, phenotypic records were removed for approximately 10% (the youngest animals) defined by hatch between the year 2018 and 2019 (reduced data set). These young animals were assigned to the validation population, and the remaining 90% animals were assigned to the training population. Thereafter, EBV and GEBV for the validation population were estimated. The predictive ability for each trait was calculated as the Pearson correlation coefficient between the EBV or GEBV estimated based on the reduced data set and the adjusted phenotype estimated from the full data set from the CUM model, while considering only animals in the validation population. Likewise, Pearson correlation coefficients between biweekly adjusted phenotypes from the full model and the biweekly EBV or GEBV estimated from the reduced data set and only considering animals in the validation population were used for the RR model.

Results

Data structure

The observed mean for FERT, HOF, and HOS was 81.27, 80.86, and 67.94%, respectively, for single records used for the CUM model as shown in Table 1. Biweekly averages of FERT, HOF, and HOS records collected during the productive life of the hen are plotted in Figure 1 and presented in Table 2. The reproductive performance trend shows an average of approximately 84.0% at the early stages of the hen's productive life, with a noticeable decline at the later stages of production/lay, decreasing to approximately 75.0% for FERT and HOF. Similarly, HOS was high at the beginning of production with an initial value of approximately 71.0%; however, after week 52, a steady decrease was observed until 62 weeks with a value of 59.0%.

Genetic parameters of different models with pedigree and genomic information

Estimates of the variance components for all the models implemented are presented in Table 4. For the CUM model using

TABLE 4 Estimates of additive, permanent, error, and phenotypic variances, heritability, and repeatability.

Trait	Model	Relationship[a]	σ_a^2	σ_{pe}^2	σ_e^2	σ_p^2	h^2	*re*
FERT	Cumulative	P	16.50	—	128.92	145.42	0.11 ± 0.02	—
		P and G	22.27	—	126.36	148.63	0.15 ± 0.02	—
	Random regression[b]	P	35.86	86.79	296.83	635.42	0.06 ± 0.01	0.19
		P and G	45.06	84.05	296.77	638.94	0.08 ± 0.01	0.20
HOF	Cumulative	P	49.21	—	157.27	206.48	0.24 ± 0.02	—
		P and G	78.67	—	144.68	223.35	0.35 ± 0.02	—
	Random regression[b]	P	108.98	117.82	477.88	799.77	0.14 ± 0.02	0.29
		P and G	137.08	109.58	478.28	815.70	0.17 ± 0.02	0.31
HOS	Cumulative	P	55.05	—	179.32	234.37	0.24 ± 0.02	—
		P and G	75.00	—	171.71	246.72	0.30 ± 0.02	—
	Random regression[b]	P	102.43	144.98	435.05	810.46	0.14 ± 0.02	0.31
		P and G	125.46	138.85	435.33	822.42	0.16 ± 0.02	0.32

[a]P, pedigree information only; P & G, pedigree and genomic information.
[b]All variance components, heritability, and repeatability for the random regression are average estimates across time point.

only pedigree information, the additive genetic variances ranged from 16.50 to 55.05 for all three traits; however, the inclusion of genomic information consequently increased the additive genetic variance by approximately 36, 60, and 36% for FERT, HOF, and HOS, respectively. Generally, the average additive genetic variances estimated using RR models were higher than those estimated using CUM and ranged from 35.86 to 108.98 for all traits with only pedigree information. With the addition of genomic information, average additive genetic variance increased to 45.06, 137.08, and 125.46 for FERT, HOF, and HOS, respectively. The residual variances from the CUM model were on average 5% lower for all traits when using pedigree and genomic information than using only pedigree information. In contrast, for the RR model, there were no substantive differences in the estimated residual variances between pedigree only and with the addition of genomic information.

An important reason for the implementation of RR is the ability to appropriately model the trajectory of longitudinal traits. This accounts for both the additive and permanent environment effect for traits with repeated records. The trends observed in the variance components estimated with the RR model are shown in Figure 2. The estimated additive genetic variance for HOF shows that the hen predominantly contributes to the observed variation, with close to zero contribution from the sire. A similar pattern was found with HOS; however, the sire contribution to the additive genetic variation increased steadily toward the end of production with a slight decline observed in the genetic variation of the hen. For FERT, both the hen and the sire contributed considerably to the additive genetic variance, with the sire having less contribution at the early stages and more contribution at the later stages of production. The permanent environment variances of the sire and hen increased gradually from the beginning of production to the end of production for all traits except for the permanent environment variance of the hen for FERT, which was constant throughout the production with a steep increase at the later production stages. In general, the addition of genomic information resulted in an increase in the additive genetic variances contributed by the hen for all traits with almost no changes in the other variance components.

Heritability estimates from the CUM model for the traits ranged from 0.11 to 0.24 for all traits with pedigree information (Tables 5, 6, 7, 8, 9, 10). On average, heritability estimates increased by approximately 28% for all traits with the addition of genomic information. Although not directly comparable to the CUM model, the average heritability estimates from the RR model with pedigree and genomic information was estimated to be 0.08, 0.17, and 0.16 for FERT, HOF, and HOS, respectively. Pedigree estimates of heritability from the RR model were an average of 23% lower than estimates with the inclusion of genomic information. The trend in heritability estimated from the RR model ranged from 0.04 to 0.10, 0.11 to 0.19, and 0.09 to 0.19 for FERT, HOF, and HOS, respectively (Figure 3). The peak for heritability estimates was observed at 42 weeks of age or from eggs produced approximately 1 month into production for all traits, and the lowest estimates were found at the end of production for FERT and HOS and 2 weeks into production for HOF. Overall, heritability estimates from pedigree and genomic information were higher than pedigree estimates.

Phenotypic and genetic correlations from the RR model are shown in Tables 5, 6, 7, 8, 9, 10. Estimated genetic correlation was found to be very high for adjacent weeks and ranged from 0.96 to 0.99. However, as the distance between the weeks increased, the correlations declined and varied from 0.57 to 0.98. In a similar pattern, phenotypic correlations were higher for closer weeks than for weeks further apart. Phenotypic correlations were substantially lower than genetic correlations and ranged from 0.02 to 0.58.

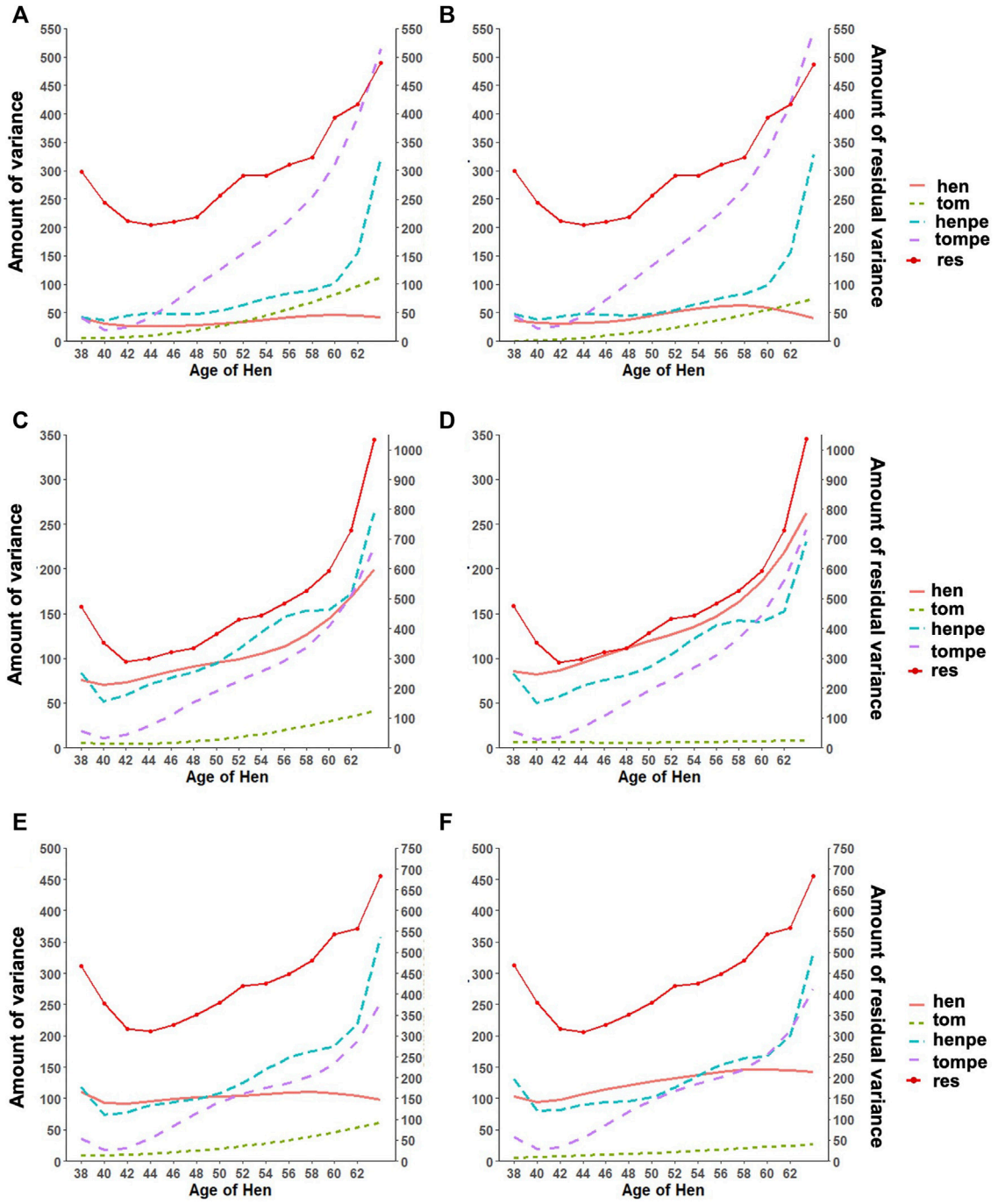


FIGURE 2
Estimates of variance components for **(A)** fertility (FERT) with pedigree, **(B)** FERT with genotypes, **(C)** hatch of fertile (HOF) with pedigree, **(D)** HOF with ssGBLUP, **(E)** hatch of set (HOS) with pedigree, and **(F)** HOS with ssGBLUP from the random regression model with linear regression for the additive genetic variance of the hen (hen) and tom (tom), the permanent environment variance of the hen (henpe) and tom (tompe), and the residual variance (res).

Predictive ability

As shown in Table 11, the predictive ability of BLUP and ssGBLUP was assessed for all models considered in this study. Across most of the studied traits and models, BLUP was outperformed by ssGBLUP with higher predictivity, and only FERT showed lower ssGBLUP predictivity than BLUP. For the CUM model, BLUP predictive ability was estimated to be 0.23,

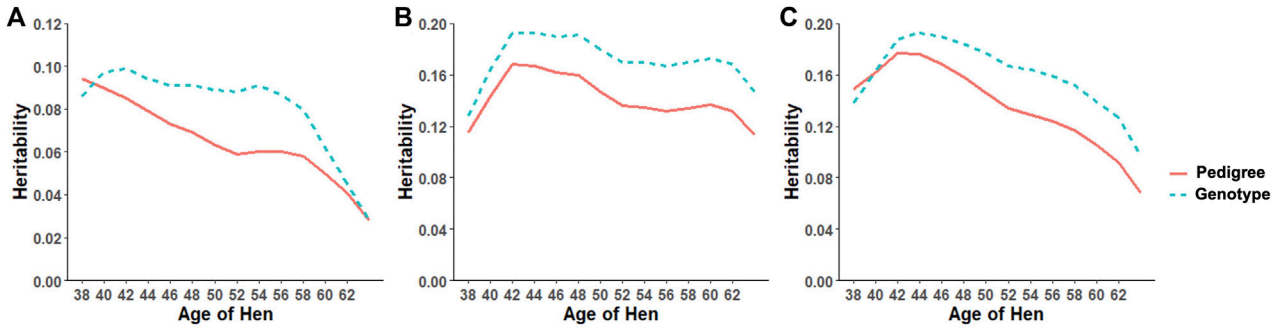


FIGURE 3
Heritability trend for different age classes for **(A)** fertility (FERT), **(B)** hatch of fertile (HOF), and **(C)** hatch of set (HOS) estimated from the random regression model using pedigree and genotypic information.

0.22, and 0.23 for FERT, HOF, and HOS, respectively. With the incorporation of genomics, these predictivities increased by 14 and 9% for HOF and HOS, respectively, and reduced by 9% for FERT. To facilitate comparison with the CUM model, average predictive ability were estimated for the RR model, which ranged from 0.11 to 0.27 for all studied traits. The biweekly trend of the predictivities estimated from the RR model is shown in Figure 4. The figure shows the maximum predictive ability estimated at approximately 44–46 weeks of age for HOF and HOS using BLUP and ssGBLUP. Conversely, the maximum predictivity for FERT was observed at 38 weeks. Predictivities estimated from BLUP were generally lower than those estimated from ssGBLUP across all biweekly records and for all traits. However, BLUP predictive ability at the early stages of production was higher than ssGBLUP for FERT.

Discussion

The present study sought to estimate genetic parameters for reproductive traits in a maternal turkey line using CUM and RR models with pedigree and genomic information. Estimated average HOF and FERT were approximately 80%, which is close to the estimates reported by Case et al. (2010). The slightly lower estimates from this study could be attributable to differences in the population used. Presently, there are no literature reports on HOS and limited reports on FERT and HOF in turkeys; hence, results from layer and broiler chickens were used for comparison. The mean HOS in this study was 18% lower than HOF, which is within the range reported for chickens (Wolc et al., 2010; Wolc et al., 2019). In accordance with previous studies, the trajectory trend in biweekly FERT, HOF, and HOS are similar to those reported in turkeys (Dunnington et al., 1990), in broilers (Heier and Jarp, 2001; Wolc et al., 2009), and in layers (Wolc et al., 2007). This trend supports the characteristically declining feature of the traits over the productive life of the hen as well as the longitudinal nature of these traits.

Genetic parameters between different models with pedigree and genomic information

The present study shows that the additive genetic effect of both the hen and sire plays a significant role in the observed variation in FERT, which is in line with the study published by Wolc et al. (2009) in broiler chickens with natural mating. Conversely, the sire additive genetic effect had a non-significant contribution to the variation observed in HOF. Similar results were reported in broiler chickens (Wolc et al., 2010), which may be due to the limited effect of the male after fertilization and more pronounced effect of the hen based on the environment provided for the developing embryo (quality of egg produced by the hen) (Wolc and Olori, 2009). Estimated additive genetic and error variances were smaller for all studied traits with the CUM model than with the RR model. The reduced variances may be because of accumulating repeated records as a single record per animal, thereby removing the covariance that exists between repeated records. With the simultaneous combination of pedigree and genomic information, additive genetic variances were higher for all the models and traits than pedigree only information. This outcome demonstrates that genomic information better captures actual relationships between individuals than the expected relationship captured by the pedigree (Hayes and Goddard, 2008). Based on the pedigree, heritability estimates for all traits ranged from 0.06 to 0.24 for all models. These estimates are within the range of 0.08–0.18 reported by Case et al. (2010). As expected, estimated heritabilities from the CUM model were higher than those from the RR model (Anang et al., 2000). This could be due to the reduced residual variance, as well as the inability to account for the correlated structure of the repeated records from the different ages. In addition, some components of the permanent environment variance could be attributed to the additive genetic variance, which would not be easily removed due

TABLE 5 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for fertility (FERT) using the pedigree random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.09 ± 0.02	0.97	0.91	0.86	0.82	0.81	0.81	0.81	0.79	0.78	0.76	0.72	0.66
40	0.25	0.09 ± 0.01	0.98	0.95	0.92	0.90	0.88	0.86	0.83	0.79	0.76	0.71	0.65
42	0.19	0.27	0.09 ± 0.01	0.99	0.97	0.96	0.93	0.89	0.84	0.80	0.75	0.69	0.64
44	0.14	0.25	0.34	0.08 ± 0.01	0.99	0.98	0.95	0.91	0.86	0.81	0.76	0.71	0.65
46	0.10	0.23	0.33	0.39	0.07 ± 0.01	0.99	0.97	0.94	0.89	0.84	0.79	0.75	0.69
48	0.08	0.20	0.31	0.39	0.44	0.07 ± 0.01	0.99	0.97	0.93	0.89	0.84	0.80	0.75
50	0.07	0.17	0.28	0.36	0.42	0.46	0.06 ± 0.01	0.99	0.97	0.93	0.90	0.86	0.81
52	0.06	0.15	0.25	0.33	0.40	0.46	0.48	0.06 ± 0.01	0.99	0.97	0.94	0.92	0.87
54	0.05	0.14	0.23	0.32	0.39	0.45	0.49	0.51	0.06 ± 0.01	0.99	0.98	0.95	0.92
56	0.05	0.12	0.21	0.29	0.37	0.43	0.47	0.50	0.54	0.06 ± 0.01	0.99	0.98	0.95
58	0.04	0.11	0.19	0.27	0.35	0.41	0.46	0.49	0.53	0.56	0.06 ± 0.01	0.99	0.97
60	0.03	0.10	0.18	0.25	0.31	0.37	0.41	0.44	0.49	0.53	0.57	0.05 ± 0.01	0.99
62	0.02	0.09	0.17	0.23	0.28	0.33	0.36	0.39	0.44	0.49	0.54	0.58	0.04 ± 0.01

TABLE 6 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for fertility (FERT) using the ssGBLUP random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.09 ± 0.02	0.99	0.98	0.97	0.95	0.92	0.88	0.85	0.83	0.81	0.81	0.83	0.87
40	0.26	0.10 ± 0.01	0.99	0.98	0.96	0.92	0.88	0.83	0.80	0.78	0.78	0.79	0.82
42	0.20	0.28	0.10 ± 0.01	0.99	0.97	0.94	0.89	0.85	0.82	0.79	0.78	0.79	0.80
44	0.15	0.26	0.34	0.09 ± 0.01	0.99	0.96	0.93	0.89	0.86	0.84	0.82	0.82	0.82
46	0.12	0.23	0.34	0.40	0.09 ± 0.01	0.99	0.97	0.94	0.92	0.89	0.87	0.86	0.85
48	0.09	0.21	0.32	0.39	0.44	0.09 ± 0.01	0.99	0.97	0.96	0.94	0.92	0.91	0.88
50	0.08	0.18	0.29	0.37	0.43	0.47	0.09 ± 0.01	0.99	0.98	0.97	0.95	0.94	0.90
52	0.07	0.16	0.26	0.34	0.41	0.46	0.49	0.09 ± 0.01	0.99	0.99	0.98	0.96	0.92
54	0.06	0.15	0.24	0.32	0.40	0.46	0.49	0.52	0.09 ± 0.01	0.99	0.99	0.97	0.93
56	0.06	0.13	0.22	0.30	0.37	0.44	0.48	0.51	0.55	0.09 ± 0.01	0.99	0.98	0.95
58	0.05	0.12	0.20	0.28	0.35	0.42	0.46	0.50	0.54	0.56	0.08 ± 0.01	0.99	0.97
60	0.04	0.11	0.18	0.25	0.32	0.37	0.42	0.45	0.49	0.53	0.57	0.06 ± 0.01	0.98
62	0.02	0.10	0.17	0.23	0.28	0.33	0.36	0.39	0.44	0.49	0.54	0.58	0.05 ± 0.01

10.3389/fgene.2022.923766

TABLE 7 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for hatch of fertile (HOF) using the pedigree random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.12 ± 0.02	0.96	0.86	0.77	0.69	0.64	0.60	0.58	0.57	0.57	0.58	0.59	0.61
40	0.25	0.14 ± 0.02	0.97	0.921	0.869	0.825	0.790	0.761	0.74	0.71	0.69	0.68	0.67
42	0.19	0.29	0.17 ± 0.02	0.98	0.95	0.92	0.89	0.87	0.83	0.79	0.76	0.73	0.70
44	0.15	0.26	0.35	0.17 ± 0.02	0.99	0.97	0.953	0.92	0.88	0.85	0.80	0.76	0.72
46	0.12	0.24	0.34	0.38	0.16 ± 0.02	0.99	0.98	0.95	0.93	0.88	0.84	0.79	0.74
48	0.10	0.22	0.32	0.37	0.39	0.16 ± 0.02	0.99	0.98	0.95	0.92	0.87	0.82	0.78
50	0.09	0.19	0.29	0.34	0.38	0.40	0.15 ± 0.01	0.99	0.98	0.95	0.91	0.86	0.82
52	0.09	0.17	0.26	0.31	0.35	0.39	0.40	0.14 ± 0.01	0.99	0.98	0.94	0.91	0.87
54	0.09	0.16	0.24	0.29	0.33	0.37	0.39	0.41	0.14 ± 0.02	0.99	0.97	0.95	0.91
56	0.10	0.15	0.21	0.26	0.31	0.35	0.38	0.41	0.43	0.13 ± 0.02	0.99	0.98	0.95
58	0.11	0.15	0.20	0.24	0.28	0.33	0.36	0.39	0.42	0.43	0.13 ± 0.02	0.99	0.98
60	0.11	0.15	0.19	0.23	0.26	0.30	0.33	0.36	0.39	0.42	0.43	0.14 ± 0.02	0.99
62	0.11	0.15	0.19	0.21	0.24	0.27	0.29	0.31	0.34	0.37	0.39	0.42	0.13 ± 0.03

TABLE 8 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for hatch of fertile (HOF) using the ssGBLUP random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.13 ± 0.02	0.96	0.84	0.80	0.73	0.68	0.65	0.63	0.63	0.64	0.67	0.69	0.73
40	0.257	0.16 ± 0.02	0.98	0.93	0.88	0.84	0.81	0.79	0.78	0.77	0.77	0.77	0.78
42	0.212	0.30	0.19 ± 0.02	0.99	0.96	0.93	0.91	0.89	0.87	0.85	0.83	0.81	0.79
44	0.163	0.28	0.36	0.19 ± 0.02	0.99	0.98	0.96	0.94	0.92	0.89	0.86	0.83	0.79
46	0.131	0.26	0.35	0.39	0.19 ± 0.02	0.99	0.98	0.97	0.95	0.92	0.88	0.84	0.80
48	0.115	0.24	0.33	0.39	0.41	0.19 ± 0.02	0.99	0.98	0.97	0.94	0.91	0.86	0.82
50	0.106	0.21	0.30	0.36	0.39	0.42	0.18 ± 0.01	0.99	0.98	0.96	0.93	0.89	0.85
52	0.104	0.19	0.27	0.33	0.37	0.40	0.42	0.17 ± 0.01	0.99	0.98	0.96	0.92	0.88
54	0.110	0.18	0.25	0.30	0.35	0.39	0.41	0.43	0.17 ± 0.02	0.99	0.98	0.95	0.92
56	0.116	0.17	0.23	0.28	0.32	0.37	0.39	0.42	0.44	0.17 ± 0.02	0.99	0.98	0.95
58	0.122	0.17	0.22	0.26	0.30	0.34	0.37	0.40	0.43	0.45	0.17 ± 0.02	0.99	0.98
60	0.126	0.17	0.21	0.25	0.28	0.32	0.34	0.37	0.41	0.43	0.44	0.17 ± 0.02	0.99
62	0.122	0.17	0.21	0.23	0.25	0.28	0.30	0.32	0.35	0.38	0.41	0.43	0.17 ± 0.02

10.3389/fgene.2022.923766

TABLE 9 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for hatch of set (HOS) using the pedigree random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.15 ± 0.02	0.95	0.84	0.74	0.67	0.63	0.60	0.58	0.57	0.56	0.56	0.55	0.55
40	0.31	0.16 ± 0.02	0.97	0.91	0.86	0.82	0.78	0.75	0.71	0.68	0.66	0.65	0.65
42	0.25	0.33	0.18 ± 0.02	0.98	0.95	0.92	0.88	0.84	0.80	0.75	0.72	0.70	0.70
44	0.19	0.30	0.39	0.18 ± 0.02	0.99	0.97	0.94	0.90	0.86	0.80	0.77	0.74	0.75
46	0.15	0.28	0.37	0.42	0.17 ± 0.02	0.99	0.97	0.94	0.90	0.85	0.81	0.79	0.79
48	0.13	0.25	0.35	0.41	0.44	0.16 ± 0.01	0.99	0.97	0.93	0.90	0.86	0.84	0.84
50	0.11	0.22	0.32	0.38	0.42	0.45	0.15 ± 0.01	0.99	0.97	0.94	0.91	0.89	0.89
52	0.11	0.20	0.29	0.35	0.39	0.43	0.45	0.13 ± 0.01	0.99	0.97	0.95	0.94	0.93
54	0.12	0.19	0.26	0.32	0.37	0.41	0.44	0.46	0.13 ± 0.01	0.99	0.98	0.97	0.96
56	0.12	0.18	0.24	0.29	0.34	0.39	0.43	0.45	0.48	0.12 ± 0.02	0.99	0.98	0.97
58	0.12	0.17	0.22	0.27	0.32	0.36	0.40	0.43	0.46	0.48	0.12 ± 0.02	0.99	0.98
60	0.12	0.16	0.21	0.25	0.29	0.33	0.36	0.39	0.43	0.45	0.47	0.11 ± 0.02	0.99
62	0.11	0.16	0.20	0.24	0.27	0.29	0.32	0.34	0.38	0.41	0.45	0.47	0.09 ± 0.02

TABLE 10 Estimates of heritability (diagonal), phenotypic correlations (below diagonal), and genetic correlations (above diagonal) for hatch of set (HOS) using the ssGBLUP random regression model.

Age	38	40	42	44	46	48	50	52	54	56	58	60	62
38	0.14 ± 0.02	0.95	0.87	0.79	0.74	0.70	0.68	0.66	0.65	0.65	0.66	0.67	0.68
40	0.32	0.16 ± 0.02	0.97	0.93	0.89	0.85	0.83	0.80	0.78	0.76	0.74	0.74	0.74
42	0.26	0.34	0.19 ± 0.02	0.98	0.96	0.94	0.91	0.88	0.85	0.82	0,79	0.78	0.78
44	0.19	0.31	0.40	0.19 ± 0.02	0.99	0.97	0.95	0.92	0.89	0.86	0.83	0.81	0.80
46	0.16	0.29	0.39	0.44	0.19 ± 0.02	0.99	0.980	0.95	0.92	0.89	0.87	0.85	0.83
48	0.14	0.26	0.37	0.42	0.45	0.18 ± 0.01	0.99	0.98	0.95	0.93	0.90	0.88	0.86
50	0.13	0.24	0.33	0.40	0.43	0.46	0.18 ± 0.01	0.99	0.98	0.96	0.93	0.91	0.89
52	0.12	0.21	0.30	0.36	0.41	0.44	0.46	0.17 ± 0.01	0.99	0.98	0.96	0.94	0.92
54	0.13	0.20	0.28	0.33	0.38	0.43	0.46	0.48	0.16 ± 0.02	0.99	0.98	0.97	0.95
56	0.13	0.19	0.26	0.31	0.35	0.40	0.44	0.47	0.49	0.16 ± 0.02	0.99	0.98	0.96
58	0.14	0.18	0.23	0.28	0.33	0.37	0.41	0.44	0.47	0.49	0.15 ± 0.02	0.99	0.98
60	0.13	0.18	0.22	0.26	0.30	0.34	0.37	0.40	0.43	0.46	0.48	0.14 ± 0.02	0.99
62	0.12	0.17	0.22	0.25	0.28	0.30	0.33	0.36	0.39	0.41	0.45	0.48	0.13 ± 0.02

10.3389/fgene.2022.923766

TABLE 11 Estimates of predictive ability using cumulative and random regression models with pedigree and genomic information.

Trait	Model	Relationship[a]	Predictive ability
FERT	Cumulative	P	0.23
		P and G	0.21
	Random regression	P	0.13
		P and G	0.11
HOF	Cumulative	P	0.22
		P and G	0.25
	Random regression	P	0.25
		P and G	0.27
HOS	Cumulative	P	0.23
		P and G	0.25
	Random regression	P	0.26
		P and G	0.27

[a]P, pedigree information only; P and G, pedigree and genomic information.

to the cumulation of records. From the RR model, heritability increased from 38 weeks of age when records were measured to a maximum at approximately week 42–46 and varied across all productive ages. Similar trends across multiple time points have been found for FERT, HOF, and egg production in turkeys (Kranis et al., 2007; Case, 2011). This indicates that RR properly accounts for environmental differences throughout the productive life of the animals. This may also indicate that different genes are being expressed at different times across the productive life of the animals. Overall, heritability estimated using the combination of pedigree and genomic relationship for all traits and models was higher than that using only pedigree relationships. Higher estimates of heritability with ssGBLUP than BLUP have also been reported in turkeys for feed conversion ratio, residual feed intake, body weight, breast meat yield, and walking score (Abdalla et al., 2019). Correlations estimated in this study showed that proximate ages had higher correlations, which declined as the ages became further apart. This pattern is similar to studies from test-day milk yield in dairy cattle (Jamrozik and Schaeffer, 1997) and goat (Brito et al., 2017). This could indicate that repeated records of the same trait collected at different ages represent different traits, especially when the time points are further apart.

Predictive ability

Predictive ability estimated based on ssGBLUP had higher estimates than BLUP for most traits. This is expected as many studies have shown lower predictivity from pedigree-based EBV than marker-based EBV (Hayes et al., 2010; Abdalla et al., 2019; Oliveira et al., 2019). Similarly, Lourenco et al. (2015) reported an increase in ssGBLUP predictive ability that ranged from 0.05 to 0.1 for birth weight in beef cattle relative to BLUP. In contrast to the higher predictions of marker-based EBV, pedigree-based EBV prediction was higher than marker-based prediction for FERT in this study. Wolc et al. (2019) found similar results for male fertility in White Leghorns layers with natural floor pen mating. This may be because fertility is influenced largely by environmental factors, and in this study, FERT had the lowest heritability among all studied traits. Another possibility is that some causal genomic regions could be located on the sex chromosomes, which were removed from these analyses. Furthermore, the CUM model had slightly higher predictability than the RR model. This is also expected because of the higher heritability estimates from the CUM model. However, these higher heritability estimates could be

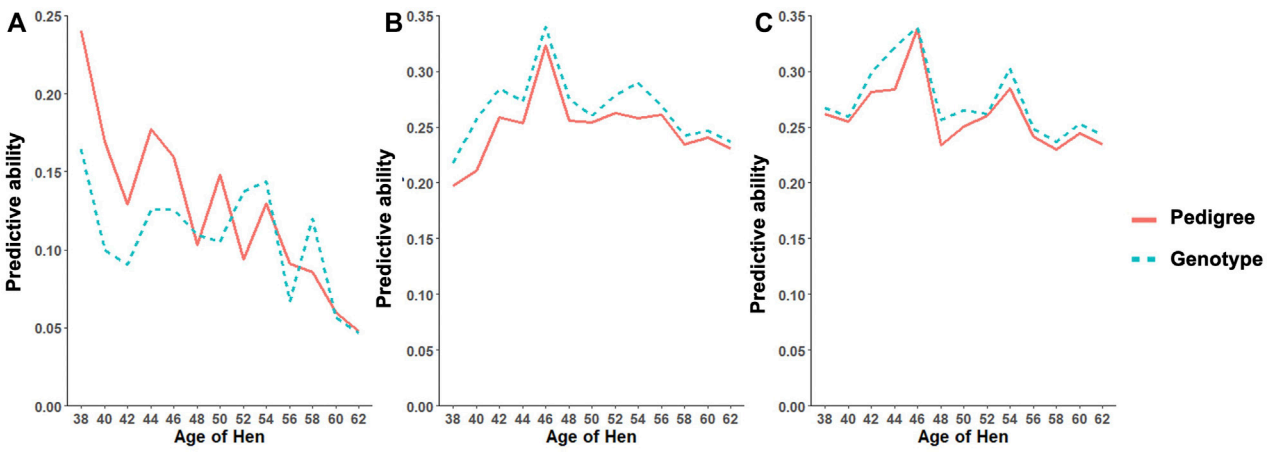


FIGURE 4
Predictive ability trend for different age classes for **(A)** fertility (FERT), **(B)** hatch of fertile (HOF), and **(C)** hatch of set (HOS) estimated from the random regression model using pedigree and genotypic information.

overestimated due to the inability to account for the permanent environment effects observed from longitudinal traits (Anang et al., 2000).

Conclusion

In this study, the applicability of RR ssGBLUP was investigated and compared to the traditional CUM model used in estimating the reproductive trait in turkeys. Our findings suggest that genomic relationships result in higher heritability estimates over traditional pedigree relationships, consequently causing higher predictive ability. In addition, the RR model captured the covariance and correlation that exist between different ages throughout the productive life of the animal. Therefore, the use of RR with the incorporation of genomic information is a feasible endeavor for analyzing longitudinal traits like FERT, HOF, and HOS in turkeys.

Data availability statement

Data that support the findings of this study are available from Hybrid Turkey upon reasonable request to the corresponding author, but restrictions apply to the availability of these data, which were used under a license of a material transfer agreement for the current study, and thus are not publicly available.

Ethics statement

The animal study was reviewed and approved by the Hendrix Genetics Animal Welfare Policy and the University of Guelph Animal Care Committee.

Author contributions

The author(s) have made the following declarations about their contributions: Conceptualization and designing the experiments: BM, EA, BW, and CB; data analysis: BM; interpretation of results: BM, EA, BW, and CB; writing and review of the manuscript: BM, EA, BW, and CB.

Funding

This study was part of the project entitled Application of genomic selection in turkeys for health, welfare, efficiency, and production traits funded by the government of Canada through the Genome Canada Genomic Application Partnership Program and administered by Ontario Genomics [recipients: BW (Industry) and CB (Academic)]. CB acknowledges support from NSERC. Apart from providing financial support, the funders have no role in the design and conduct of the studies. Researchers maintain independence in conducting their studies and report the outcomes regardless of the results. The decision to publish the findings rests solely with the researchers.

Acknowledgments

The authors extend their gratitude to the managers and personnel of the Hybrid Turkeys pedigree farm (Kitchener, Canada) for collecting and providing data used in this study.

Conflict of interest

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

Publisher's note

References

Abdalla, E. E. A., Schenkel, F. S., Emamgholi Begli, H., Willems, O. W., van As, P., Vanderhout, R., et al. (2019). Single-step methodology for genomic evaluation in turkeys (*Meleagris gallopavo*). *Front. Genet.* 10, 1248. doi:10.3389/fgene.2019.01248

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

Akaike, H. (1998). *Information theory and an extension of the maximum likelihood principle*. New York, NY: Springer, 199–213. doi:10.1007/978-1-4612-1694-0_15

Anang, A., Mielenz, N., and Schüler, L. (2000). Genetic and phenotypic parameters for monthly egg production in white leghorn hens. *J. Anim. Breed. Genet.* 117, 407–415. doi:10.1046/j.1439-0388.2000.00258.x

Anang, A., Mielenz, N., and Schüler, L. (2002). Monthly model for genetic evaluation of laying hens II. Random regression. *Br. Poult. Sci.* 43, 384–390. doi:10.1080/00071660120103657

Brito, L. F., Silva, F. G., Oliveira, H. R., Souza, N. O., Caetano, G. C., Costa, E. V., et al. (2017). Modelling lactation curves of dairy goats by fitting random regression models using legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98, 73. doi:10.1139/cjas-2017-0019

Case, L. A. (2011). Improving the efficiency of Turkey breeding programs through selection index design, technological advancements, and management optimization. Theses & Dissertations. University of Guelph.

Case, L. A., Kelly, M. J., Miller, S. P., and Wood, B. J. (2010). Genotype x environment interaction as it relates to egg production in turkeys (*Meleagris gallopavo*). *J. Anim. Sci.* 88, 1957–1966. doi:10.2527/jas.2009-2004

Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42, 2. doi:10.1186/1297-9686-42-2

Dunnington, E. A., Van Krey, H. P., Hulet, R. M., and Denbow, D. M. (1990). Genetic influences on seasonal decline in the fertility of female turkeys. *Poult. Sci.* 69, 365–368. doi:10.3382/ps.0690365

Emamgholi Begli, H., Schaeffer, L. R., Abdalla, E., Lozada-Soto, E. A., Harlander-Matauschek, A., Wood, B. J., et al. (2021). Genetic analysis of egg production traits in turkeys (*Meleagris gallopavo*) using a single-step genomic random regression model. *Genet. Sel. Evol.* 53, 61. doi:10.1186/s12711-021-00655-w

Englishby, T. M., Banos, G., Moore, K. L., Coffey, M. P., Evans, R. D., and Berry, D. P. (2016). Genetic analysis of carcass traits in beef cattle using random regression models. *J. Anim. Sci.* 94, 1354–1364. doi:10.2527/jas.2015-0246

FAO (2021). FAOSTAT. Available at: http://www.fao.org/faostat/en/#data/QL/visualize (Accessed February 2, 2021).

Hayes, B. J., and Goddard, M. E. (2008). Technical note: prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86, 2089–2092. doi:10.2527/jas.2007-0733

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., and Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. *PLoS Genet.* 6, 1001139. doi:10.1371/journal.pgen.1001139

Heier, B. T., and Jarp, J. (2001). An epidemiological study of the hatchability in broiler breeder flocks. *Poult. Sci.* 80, 1132–1138. doi:10.1093/ps/80.8.1132

Jamrozik, J., and Schaeffer, L. R. (1997). Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J. Dairy Sci.* 80, 762–770. doi:10.3168/jds.S0022-0302(97)75996-4

Kheirabadi, K., and Rashidi, A. (2016). Genetic description of growth traits in markhoz goat using random regression models. *Small Ruminant Res.* 144, 305–312. doi:10.1016/j.smallrumres.2016.10.003

Kranis, A., Su, G., Sorensen, D., and Woolliams, J. A. (2007). The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poult. Sci.* 86, 470–475. doi:10.1093/ps/86.3.470

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi:10.3168/jds.2009-2061

Lourenco, D. A. L., Tsuruta, S., Fragomeni, B. O., Masuda, Y., Aguilar, I., Legarra, A., et al. (2015). Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *J. Anim. Sci.* 93, 2653–2662. doi:10.2527/JAS.2014-8836

Makanjuola, B. O., Olori, V. E., and Mrode, R. A. (2021). Modeling genetic components of hatch of fertile in broiler breeders. *Poult. Sci.* 100, 101062. doi:10.1016/j.psj.2021.101062

Meyer, K. (2007). WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8, 815–821. doi:10.1631/jzus.2007.B0815

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

Nestor, K. E., Noble, D. O., Zhu, J., and Moritsu, Y. (1996). Direct and correlated responses to long-term selection for increased body weight and egg production in turkeys. *Poult. Sci.* 75, 1180–1191. doi:10.3382/ps.0751180

Oliveira, H. R., Lourenco, D. A. L., Masuda, Y., Misztal, I., Tsuruta, S., Jamrozik, J., et al. (2019). Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *J. Dairy Sci.* 102, 2365–2377. doi:10.3168/jds.2018-15466

Rovadoscki, G. A., Petrini, J., Ramirez-Diaz, J., Pertile, S. F. N., Pertille, F., Salvian, M., et al. (2016). Genetic parameters for growth characteristics of free-range chickens under univariate random regression models. *Poult. Sci.* 95, 1989–1998. doi:10.3382/ps/pew167

Sasaki, O., Aihara, M., Nishiura, A., Takeda, H., and Satoh, M. (2015). Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. *J. Dairy Sci.* 98, 5781–5795. doi:10.3168/jds.2014-9152

Schaeffer, L. R., and Dekkers, J. C. M. (1994). "Random regressions in animal models for test-day production in dairy cattle," in Proceedings of the 5th world congress on genetics applied to livestock production, Guelph, ON, Canada, August 7–12, 1994, 443–446.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Whalen, A., and Hickey, J. M. (2020). AlphaImpute2: fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. bioRxiv. doi:10.1101/2020.09.16.299677

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., and Dekkers, J. C. M. (2019). Genetics of male reproductive performance in white leghorns. *Poult. Sci.* 98, 2729–2733. doi:10.3382/ps/pez077

Wolc, A., Lisowski, M., and Szwaczkowski, T. (2007). Genetic evaluation of laying hens under fixed regression animal models. *Arch. Anim. Breed.* 50, 279–287. doi:10.5194/aab-50-279-2007

Wolc, A., and Olori, V. E. (2009). "Genetics of hatchability-egg quality from the perspective of a chick," in Proceedings of 6th european poultry genetic symposium, Będlewo, Poland, September 30–October 2, 2009, 10.

Wolc, A., White, I. M. S., Hill, W. G., and Olori, V. E. (2010). Inheritance of hatchability in broiler chickens and its relationship to egg quality traits. *Poult. Sci.* 89, 2334–2340. doi:10.3382/ps.2009-00614

Wolc, A., White, I. M. S., Olori, V. E., and Hill, W. G. (2009). Inheritance of fertility in broiler chickens. *Genet. Sel. Evol.* 41, 47. doi:10.1186/1297-9686-41-47

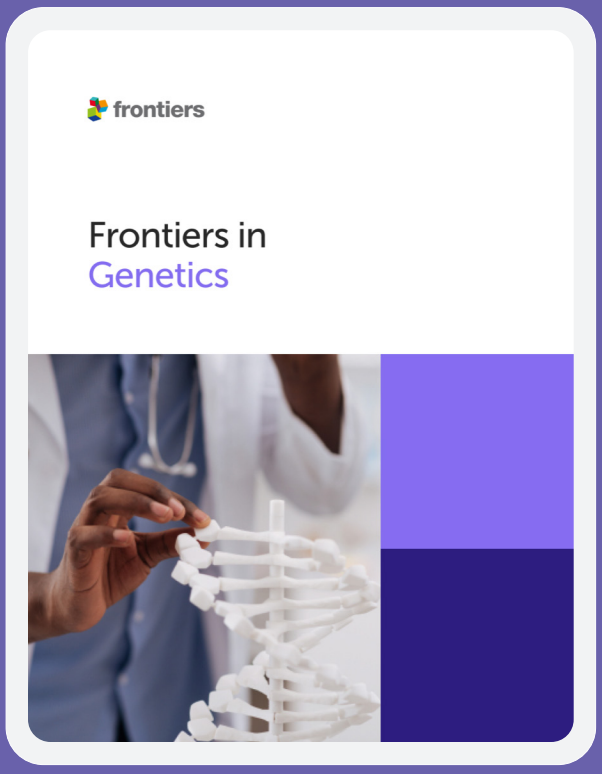
Frontiers in Genetics

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