



OPEN ACCESS

EDITED BY

Maslyn Greene,
Clemson University, United States

REVIEWED BY

Henrique A. Mulim,
Purdue University, United States
Xiaolong Kang,
Ningxia University, China
Enliang Song,
Shandong Academy of Agricultural Sciences,
China

*CORRESPONDENCE

Pengjia Bao
✉ baopengjia@caas.cn
Ping Yan
✉ pingyanlz@163.com

RECEIVED 05 December 2024

ACCEPTED 21 January 2025

PUBLISHED 24 June 2025

CITATION

Zhang M, Huang C, Bao Q, Wang T, Ma X,
Meng G, Zhou X, Zheng Q, Chu M, Liang C,
Guo X, Bao P and Yan P (2025)

Whole-genome resequencing revealed the
genetic diversity and body weight-related
genes of Gannan yak.

Front. Vet. Sci. 12:1540523.

doi: 10.3389/fvets.2025.1540523

COPYRIGHT

© 2025 Zhang, Huang, Bao, Wang, Ma, Meng,
Zhou, Zheng, Chu, Liang, Guo, Bao and Yan.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Whole-genome resequencing revealed the genetic diversity and body weight-related genes of Gannan yak

Minghao Zhang^{1,2}, Chun Huang^{1,2}, Qi Bao³, Tong Wang^{1,2},
Xiaoming Ma^{1,2}, Guangyao Meng^{1,2}, Xuelan Zhou^{1,2},
Qingbo Zheng^{1,2}, Min Chu^{1,2}, Chunnian Liang^{1,2}, Xian Guo^{1,2},
Pengjia Bao^{1,2*} and Ping Yan^{1,2*}

¹Key Laboratory of Animal Genetics and Breeding on Tibetan Plateau, Ministry of Agriculture and Rural Affairs, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, China, ²Key Laboratory of Yak Breeding in Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, China, ³Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

Body weight (BW) is a crucial indicator of animal growth and development, significantly influencing animal husbandry practices. Previous research has identified several genes associated with BW in certain yak breeds. However, the genetic basis of BW in Gannan yaks has not been reported. In this study, 309 yaks from six breeds across five provinces in China were sampled. This collection included 247 54-month-old female Gannan yaks, along with 20 Xizang yaks, 15 Muli yaks, 10 Pamir yaks, 9 Bazhou yaks, and 8 Zhongdian yaks. Body weight measurements were recorded for the Gannan yaks. Initial analyses of runs of homozygosity (ROH), nucleotide diversity, and linkage disequilibrium (LD) decay among the six yak breeds revealed that Gannan yaks exhibited the lowest ROH, the highest nucleotide diversity, and the fastest LD decay, indicating rich genetic diversity. Subsequently, a genome-wide association study (GWAS) identified 19 BW-related genes in the Gannan yaks, with *PMAIP1*, *GABBR1*, *LRPPRC*, and *PPP1R11* identified as key genes. Genome-wide scanning of Group 1 and Group 2 (detailed in section 2.4) identified 90 genes, and Gene Ontology (GO) analysis highlighted *FGF2*, *SHH*, and *WNT11* as significantly associated with growth, development, and metabolism. Three overlapping genes were identified between GWAS and genome-wide scans. Further analyses, including nucleotide diversity, LD analysis of significant GWAS sites, allele frequency analysis, and SNP association studies, suggested that *DRC1* and *SELENOI* are novel candidate genes for BW in Gannan yaks. These findings provide a molecular foundation for the genetic improvement of Gannan yaks.

KEYWORDS

Gannan yak, body weight, genetic diversity, genome-wide association study, genome-wide scanning

1 Introduction

Gannan yak, an ancient breed native to the Gannan Tibetan Autonomous Prefecture, is well-adapted to plateau environments and plays a vital role in the local economy (1). These yaks are known for their rich content of amino acids and trace elements in meat, highlighting their nutritional value. However, due to reliance on natural grassland grazing and the lack of

scientific feeding and breeding practices, the breed has experienced degradation and decreased production performance. This has led to significant individual variations in BW among animals of the same age and gender.

Body weight is a critical growth trait directly reflecting the production performance of yaks and serving as an important indicator for meat production potential (2). Enhancing BW research is a key strategy to improve the productivity of Gannan yaks. Identifying essential functional genes associated with BW can significantly aid in the breeding and improvement of yak production performance. However, traditional crossbreeding efforts between yaks and cattle face challenges such as male sterility, chromosomal incompatibility, and genomic conflicts, limiting their effectiveness in improving yak breeds (3). Advances in DNA sequencing technology provide valuable tools for strengthening breeding efficiency (4). Since the first yak reference genome was sequenced in 2012 (5) and a chromosome-scale genome was published in 2020, researchers have gained new opportunities to conduct large-scale molecular breeding (6). Previous studies have identified BW-associated genes such as *MFSD4*, *LRR37B*, and *NCAM2* in Maiwa yak through GWAS (2), and genes like *AHR*, *HPGDS*, *HSF1*, *SOX5*, *SOX6*, and *SOX8* in Ashidan yak through copy number variation analysis (CNV) (7–11). However, compared to livestock such as cattle, pig, and sheep, the breeding progress of yak has lagged behind. The Animal Quantitative Trait Loci (QTL) Database (Animal QTLdb) is a comprehensive repository that collects publicly available data on QTL, including phenotype and expression QTL (eQTL), GWAS, and CNV for livestock species. Among the extensively studied species, cattle, pigs, and sheep dominate, with 192,336 QTL records for cattle, 56,615 for pigs, and 5,058 for sheep, representing hundreds of base traits and trait variants curated from thousands of publications. In contrast, research on yak, particularly the Gannan yak, has significantly lagged behind in terms of QTL data and breeding advancements. Despite its economic and ecological importance in high-altitude regions, the breeding progress of Gannan yak remains limited. There is an urgent need to leverage molecular breeding techniques to identify key molecular markers associated with its growth traits.

In this study, we utilized a combination of GWAS, genome-wide scanning, and SNP association analyses to identify candidate genes associated with the BW trait in Gannan yaks. Such efforts would accelerate artificial selection, enhance productivity, and help preserve this valuable breed while bridging the gap in genomic research between yak and other livestock species.

2 Materials and methods

2.1 Sample collection and phenotypic measurement of Gannan yaks

A total of 309 pasture-raised yaks were collected, all of them were adult domesticated yaks. They include 9 Bazhou yaks and 10 Pamir yaks, which were distributed in Xinjiang Uygur Autonomous Region of China. Twenty yaks in the Tibet Tibetan Autonomous Region of China, latter text is simply called Xizang yak; 15 Muli yaks in Sichuan Province, China; 8 Zhongdian yaks in Yunnan Province and 247 54-month-old female Gannan yaks in Gansu Province, China (Figure 1). Blood samples of 5 mL were obtained from the veins of each yak, transferred into EDTA anticoagulant tubes, and preserved at -20°C until subsequent analysis. The BW data (Table 1) of 247 Gannan yaks were collected in 54-month-olds in October 2023, and descriptive statistics were conducted using R (v4.3.3) language. All yak breeds not have pedigree records. The body weight of Gannan yaks was measured using a floor scale in the morning before fasting. The measurements were taken on two consecutive days, and the average value was calculated, expressed in kilograms (kg).

2.2 Genomic DNA extraction, resequencing and SNP detection

DNA was extracted from blood samples using the EasyPure Blood Genome DNA Kit (Quanjin, Beijing). Concentration was measured with a Qubit fluorometer (Invitrogen, United States), and integrity was

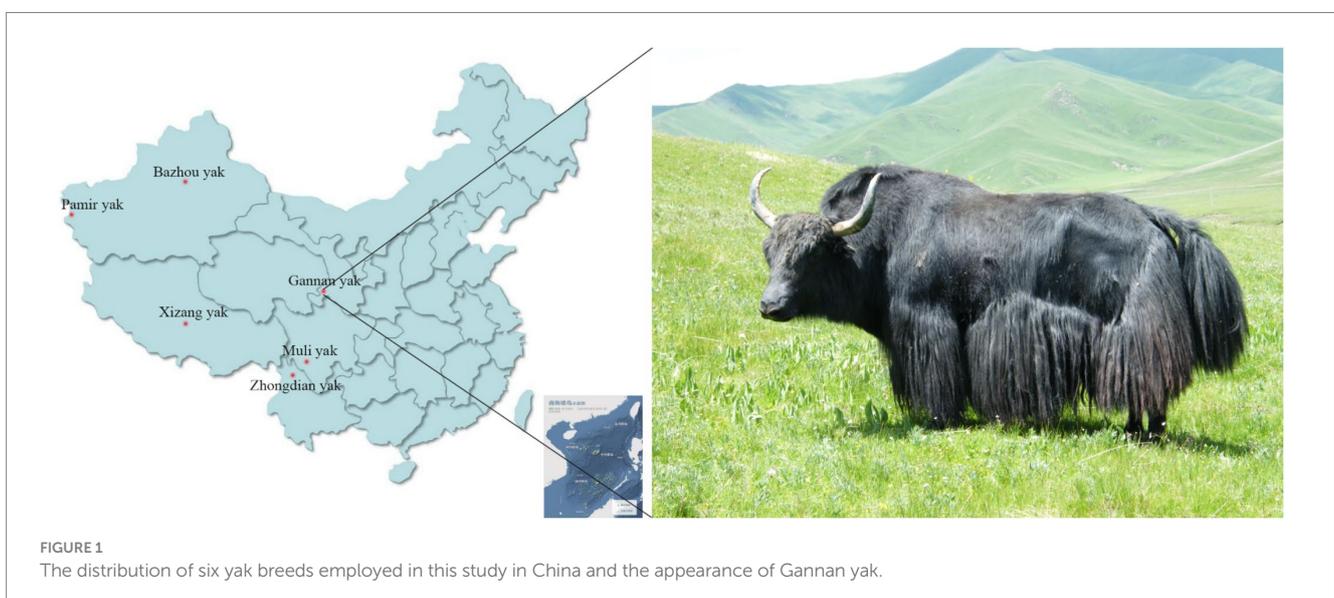


TABLE 1 Descriptive statistics of BW of Gannan yaks.

Item	Number	Mean	SD	Max	Min	SE
All	247	218.0	34.6	337.1	162.2	2.2
Group 1	20	287.0	16.6	337.1	270.3	3.7
Group 2	20	171.5	3.8	175.4	162.2	0.8

Mean, mean value; SD, standard deviation; Max, maximum value; Min, minimum value; SE, standard error; Group 1—the 20 of the largest body weight yaks in Gannan yak; Group 2—the 20 of the smallest weight yaks in Gannan yak.

checked via agarose gel electrophoresis. DNA was fragmented to 300–500 bp, followed by terminal repair, 3'-adenylation, adapter ligation, and PCR amplification. The single-stranded PCR product was cyclized, uncyclized DNA was digested, and rolling cycle amplification produced DNA nanoballs (DNB). DNBs were loaded onto nanoarrays and sequenced on the DNBSEQ-T7 platform (BGI, China). Sequencing reads were stored in FASTQ format after base calling analysis. The reads were mapped to the yak reference genome (*Bos grunniens.LU_Bosgru_v3.0*) from Ensemble using the Sentieon (12) software (V202112). Sentieon was used to detect the mutation sites of each sample to obtain the gVCF of each sample. Sentieon was used for joint-calling, and the gVCF of all samples was analyzed jointly to obtain the variation results of each individual in the population. To ensure the accuracy of SNP, preliminary hard filtering was performed on the SNP loci obtained after joint analysis (SNP hard filtering criteria: "QD 60.0 | | MQ 3.0 | | MQRankSum <-12.5 | | ReadPosRankSum <-8.0"). To prevent the influence of gender on subsequent analysis and reduce the SNP call errors caused by error mapping or insertion and deletion (InDels), VCFtools (v0.1.16) (13) was used to extract 1–29 autosomes. Only high-quality SNPs (coverage depth ≥ 3 , RMS mapping quality ≥ 20 , maf ≥ 0.05 , missed ≤ 0.2) were selected for subsequent analysis.

2.3 Population structure and genetic analysis

A principal component analysis (PCA) and runs of homozygosity analysis were conducted using PLINK (v1.07) (14). ROH were identified for estimating homozygosity using PLINK (--homozyg-density 50 --homozyg-gap 100 --homozyg-kb 300 --homozyg-snp 50 --homozyg-window-het 3 --homozyg-window-snp 50 --homozyg-window-threshold 0.05). VCFtools was used to estimate nucleotide diversity in each group with the parameter "--window-pi 50000 --window-pi-step 5000." Linkage disequilibrium decay with physical distance between SNPs was calculated and visualized by using PopLDdecay (v3.42) (15). The range of gene annotation was determined based on the r^2 value of LD decay ($r^2 \approx 0.1$) of 247 Gannan yaks.

2.4 Genome-wide association study

The FarmCPU model implemented in the rMVP (v1.2.0) (16) package of the R was utilized to identify SNPs significantly associated with BW phenotypic traits. Quality control was performed on the SNP data, and the genome-wide significance threshold was set to $1/N_{\text{snp}}$, where N_{snp} represents the total

number of SNPs remaining after quality control. Correction for multiple testing was performed using the Bonferroni method. FarmCPU is a mixed linear model-based approach that alternates between fixed and random effects to balance the detection of true associations and control for confounding factors. In this study, PCA was incorporated as fixed effects to adjust for population stratification, while kinship, calculated automatically by rMVP, was included as random effects to account for genetic relatedness among individuals.

The model is expressed as follows:

$$y_i = Mi1b1 + Mi2b2 + \dots + Mi1b1 + Mi2b2, y_i = Mi1b1 + Mi2b2 + \dots + Minbn + Zijuj + ei \quad (1)$$

$$y_i = Vi + eiyi = Vi + ei \quad (2)$$

In Equations 1, 2, Equation 1 is a fixed-effect model, Equation 2 is a random-effects model, y_i represents the phenotypic observation of the i -th individual, Min represents the classification result of a total of n potential correlation sites included in the model, bn indicates the effect value corresponding to the site, Zij indicates the classification result of the j -th marker of the i -th individual, uj is the effect value of Zij , Vi represents the total genetic effect of the i -th individual, and ei is the residual vector, subject to $e \sim N(0, I\sigma^2)e \sim N0, I\sigma^2$. When executing the FarmCPU model, styles Equations 1, 2 are alternate operations. The alternating operation between Equation 1 (fixed-effect model) and Equation 2 (random-effect model) ensures the detection of SNPs associated with phenotypic traits while controlling for population structure and genetic background. By incorporating PCA and kinship, the FarmCPU model effectively minimizes false positives and improves the accuracy of GWAS results.

2.5 Genome-wide scanning

Twenty individuals from 247 54-month-old female Gannan yaks with the largest BW were grouped into Group 1, and the 20 individuals also from 247 54-month-old female Gannan yaks with the smallest BW were grouped into Group 2. A t -test was conducted to assess whether there was a significant difference in BW between the two groups. We performed genome-wide scanning of Group 1 and Group 2 using F_{st} and θ_{π} -ratio. F_{st} values and θ_{π} ratios were calculated using 50-kb windows with 5-kb sliding steps in VCFtools. For both metrics, windows in the top 1% were deemed significant. Regions where top 1% windows of F_{st} and θ_{π} -ratio overlapped were identified as candidate selection regions.

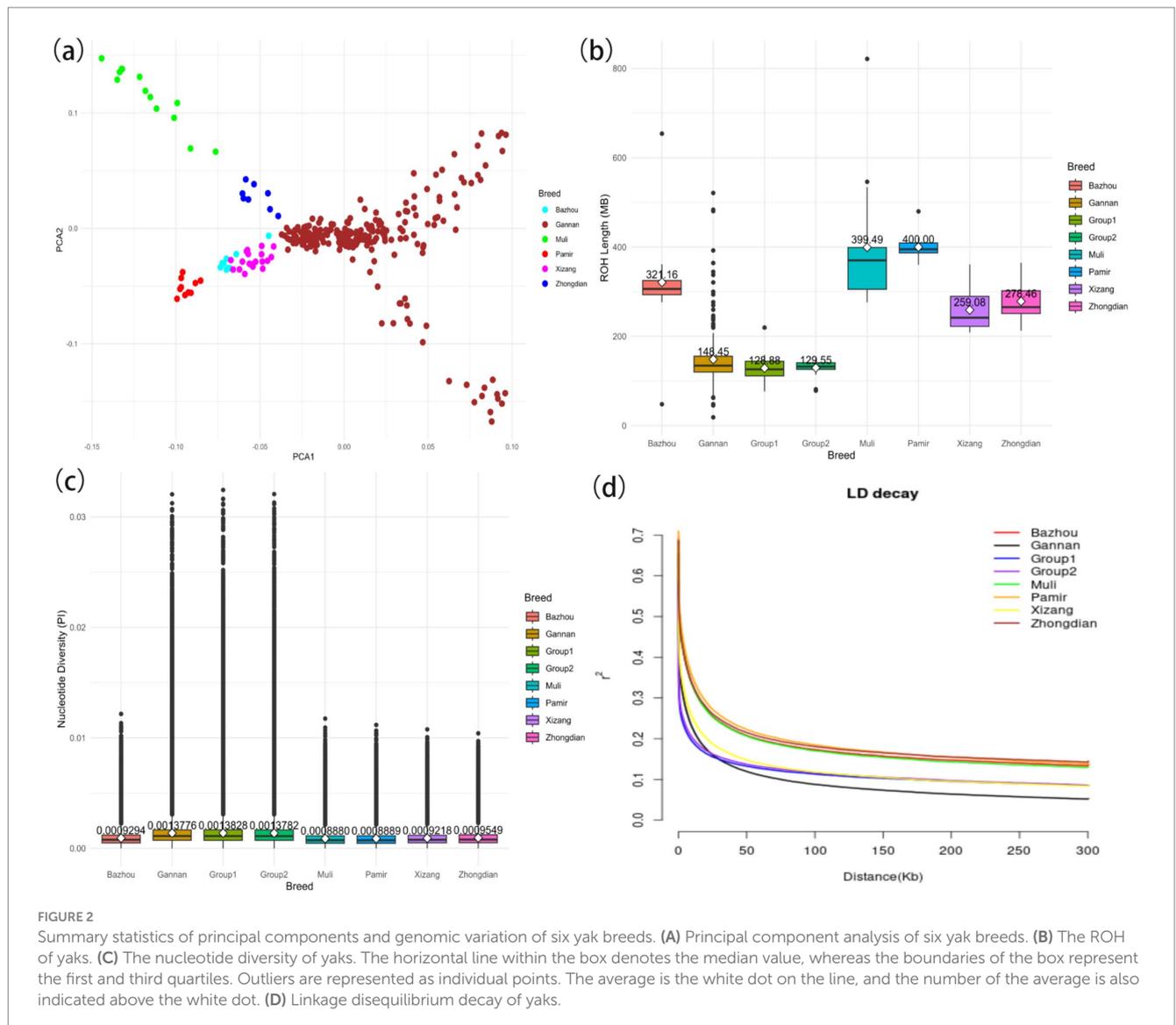


FIGURE 2

Summary statistics of principal components and genomic variation of six yak breeds. (A) Principal component analysis of six yak breeds. (B) The ROH of yaks. (C) The nucleotide diversity of yaks. The horizontal line within the box denotes the median value, whereas the boundaries of the box represent the first and third quartiles. Outliers are represented as individual points. The average is the white dot on the line, and the number of the average is also indicated above the white dot. (D) Linkage disequilibrium decay of yaks.

2.6 Candidate gene analysis

In this study, the yak 3.0 reference genome (*Bos grunniens*. LU_Bosgru_v3.0) in Ensembl was used for gene annotation. The gene annotation software was BEDTools (v2.31.1) (17). The range of genes annotated for the significant GWAS site is determined based on the LD decay r^2 value ($r^2 \approx 0.1$) in Gannan yaks. The genes for the genome-wide scan are annotated based on the 1% common window range. Venny 2.1.0¹ was used for the Venn diagram. Metascape² was used for the biological function analyses. The VCFtools was used to calculate the nucleotide diversity in the target region with 5-kb windows and to extract haplotypes. The PLINK software was used to calculate the LD.

1 <https://bioinfogp.cnb.csic.es/tools/venny/index.html>

2 <https://metascape.org/gp/index.html>

3 Results

3.1 Genetic diversity analysis of yak

This study examined the genetic diversity of 309 yaks from six distinct breeds. Principal component analysis (PCA) revealed clear genetic differentiation between Gannan yaks and the other five breeds (Figure 2A). Notably, Gannan yaks exhibited shorter runs of homozygosity (ROH) segments compared to Pamir yaks, indicating lower recent inbreeding levels (Figure 2B). Additionally, the ROH of Group 1 yaks was shorter than that of Group 2 yaks, suggesting a greater genetic diversity in the former. Nucleotide diversity analyses highlighted that Gannan yaks had the highest diversity among the breeds, contrasting with the lowest values observed in Muli yaks (Figure 2C). Furthermore, Group 1 demonstrated higher nucleotide diversity compared to Group 2, reflecting greater genetic variation in larger individuals. Linkage disequilibrium (LD) decay analysis showed that Gannan yaks exhibited the fastest decay rates, indicating a higher recombination

rate or older population structure (Figure 2D). These findings collectively underscore the rich genetic diversity and unique evolutionary history of Gannan yaks, with Group 1 displaying greater variation than Group 2.

3.2 Statistics of Gannan yak body weight

BW data for 247 Gannan yaks were summarized (Table 1). The overall mean BW was 218.0 ± 34.6 kg, with Group 1 exhibiting significantly higher BW (287.0 ± 16.6 kg) compared to Group 2 (171.5 ± 3.8 kg) (Figure 3). The large standard deviation and range reflect the extensive phenotypic variation within the population. This significant BW difference between the groups provided a basis for identifying genetic factors underlying this trait.

3.3 Genotype calling and analyzing of 247 Gannan yaks

After filtering, 12411513 valid high-quality SNPs of 247 Gannan yaks were retained. The PCA1 and PCA2 divided 247 Gannan yaks into three groups in Figure 4A, demonstrating stratification among the Gannan yaks. As a result, PCA = 3 was incorporated into the FarmCPU model for adjustment, and the kinship file generated during the rMVP analysis was also included in the FarmCPU model to

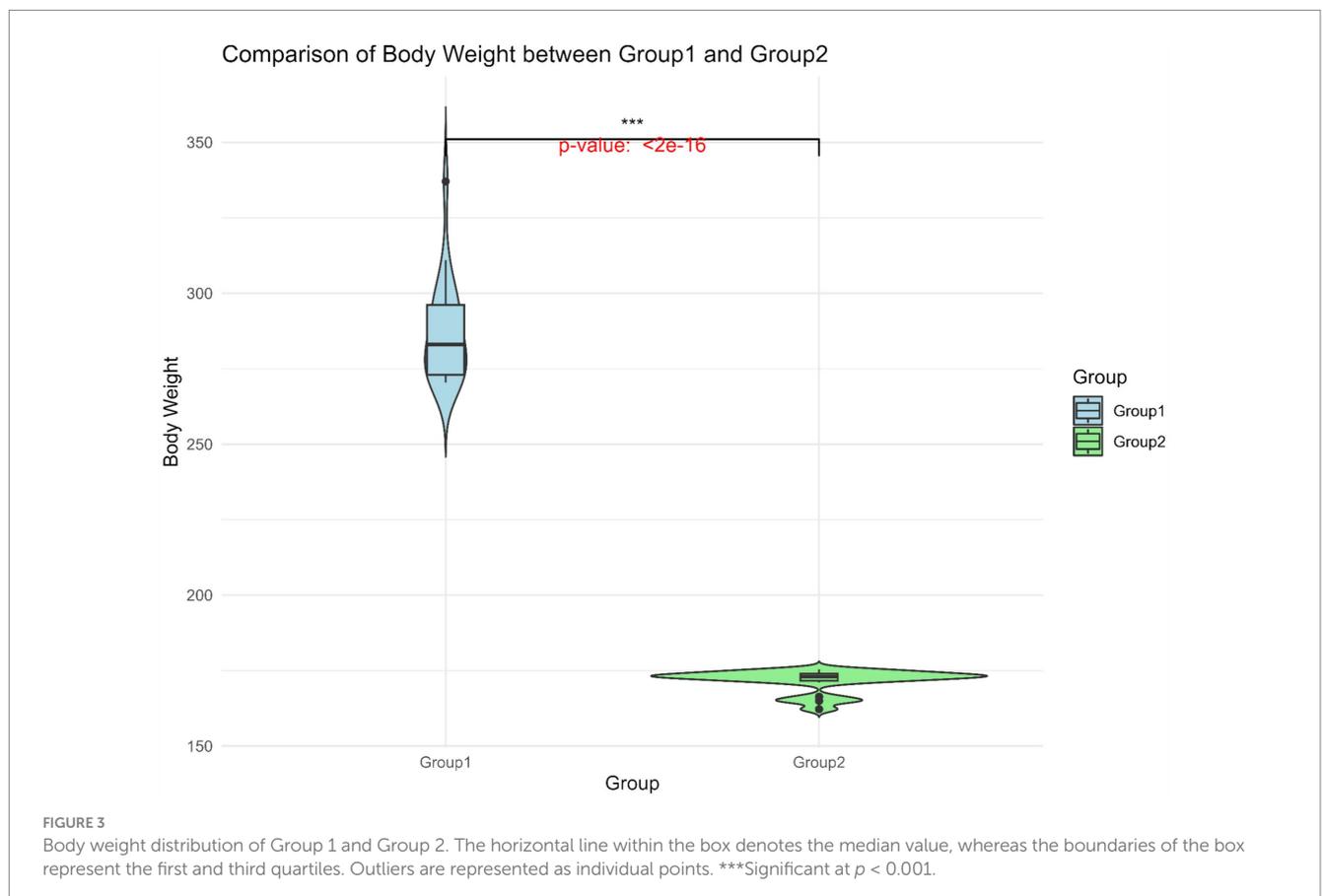
improve the accuracy of GWAS. LD analysis revealed that LD decayed to a stable level ($r^2 \approx 0.1$) within 50–60 kb, indicating that regions within this range were suitable for candidate gene annotation (Figure 4B). Therefore, we selected a ± 50 kb window around significant GWAS site as the annotation range for candidate genes.

3.4 GWAS analysis results

Using the FarmCPU model, 13 significant SNPs associated with BW were identified (Figures 5A,B). These SNPs annotated 19 genes (Table 2), including several with known roles in growth and metabolism. For example, *LRPPRC* is involved in cytoskeletal regulation and bone development, *PMAIP1* is associated with obesity, and *PPP1R11* has roles in cell proliferation and fat metabolism. Additionally, *GABBR1*, involved in fatty acid metabolism, emerged as a novel candidate. These genes provide insights into the molecular mechanisms influencing BW in Gannan yaks.

3.5 The result of genome-wide scanning

From the results of Figures 2, 3, the genotype and phenotype of Group 1 are different from Group 2. To further explore the genes associated with BW, we searched the whole genomic region of Group 1 and Group 2 using two statistical methods (F_{st} and θ_{π} -ratio) to compare the Group 1 and Group 2 based on sliding 50-kb windows



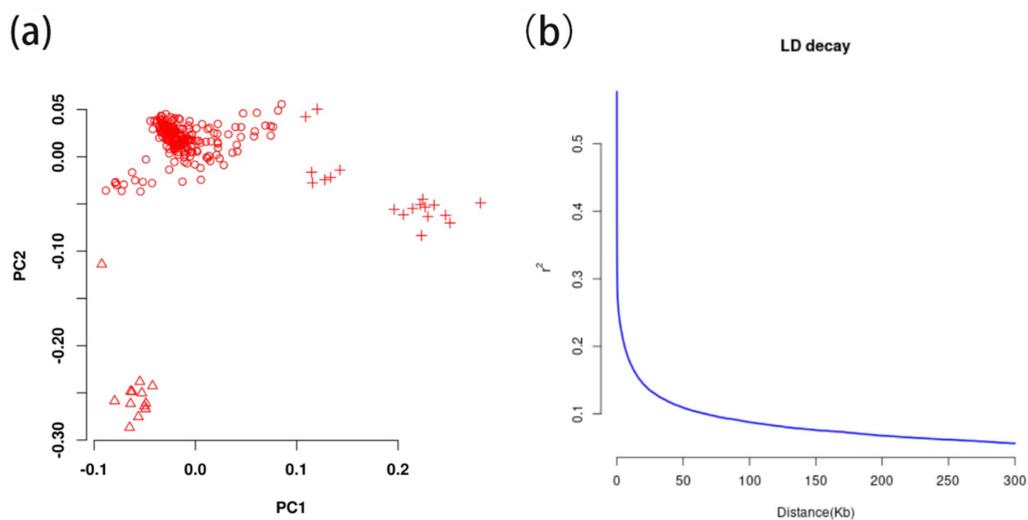


FIGURE 4

The principal component analysis and linkage disequilibrium decay calculation of 247 Gannan yak. (A) Principal components analysis of 247 Gannan yaks. Different symbols represent the grouping situation. (B) Linkage disequilibrium decay of 247 Gannan yaks.

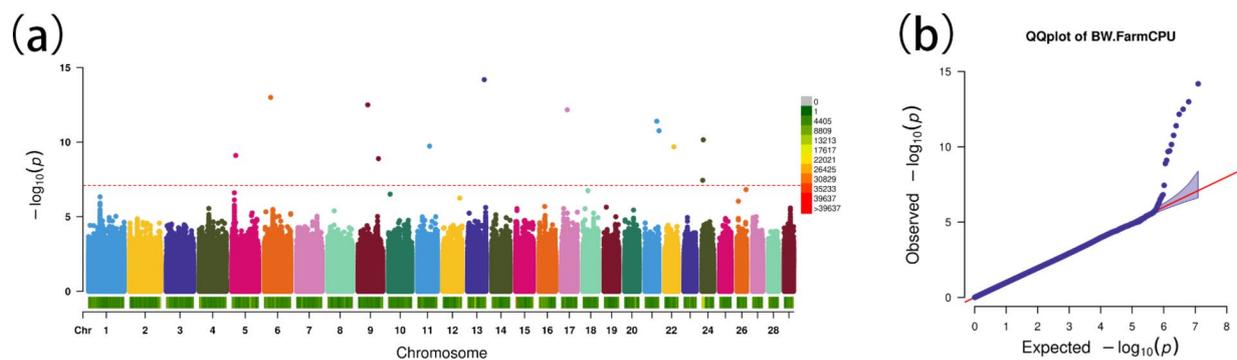


FIGURE 5

Genome-wide association study for body weight in Gannan yaks. (A) Manhattan map of genome-wide association study. (B) Quantile-Quantile plots from the genome-wide association study.

with 5-kb steps (Figures 6A,B). The regions within the top 1% windows were considered candidate genomic regions. We identified 489 and 485 candidate genes through F_{st} and θ_{π} -ratio (Figure 6C). On combining F_{st} and θ_{π} -ratio analyses, we identified 90 overlapping genes (Figures 6C, 7A). To enhance the understanding of the functions associated with overlapping genes, an enrichment analysis was conducted using Metascape. Among them, *FGF2*, *SHH* and *WNT11* were enriched in five Gene Ontology biological processes: growth, positive regulation of biological process, regulation of biological process, developmental process and metabolic process (Figure 7B). *OTOF*, *DRC1*, and *SELENOI* overlap in the GWAS and the genome-wide scanning, prompting more comprehensive future studies.

3.6 Gene analysis

To further investigate the association between the overlapping genes *OTOF*, *DRC1*, and *SELENOI* and BW in Gannan yaks,

we calculated the nucleotide diversity and nucleotide diversity ratio between Group 1 and Group 2 in *Chr9* 97665582–97797598 regions (Figures 8A,B). In *DRC1* and *SELENOI*, the nucleotide diversity in most windows of Group 1 was higher than that of Group 2. After that, we calculated the LD values of the significant GWAS site (*Chr9* 97711609) in the *Chr9* 97665582–97797598 region (Figure 8C), the results showed that there was no $R^2 \geq 0.8$ SNPs were located in the *OTOF* gene, meanwhile, the *Chr9* 97763380 ($R^2 > 0.8$) site fell inside the *SELENOI* gene. It indicated that *DRC1* and *SELENOI* were more likely to be the candidate genes for BW of Gannan yak. To further prove that *DRC1* and *SELENOI* are candidate genes for BW, we selected two sites (*Chr9* 97711609 and *Chr9* 97763380) for analysis. The allele frequency showed that most of the mutant-type (*Chr9* 97711609 for T, *Chr9* 97763380 for C) of the two sites were concentrated in Group 1, while most of the wild-type (*Chr9* 97711609 for G, *Chr9* 97763380 for T) were concentrated in Group 2 (Figures 8D,E). This can also prove to a certain extent that the variation of Group 1 is richer than that of Group 2, and also shows

TABLE 2 The information on candidate genes from the genome-wide association study.

Chromosome	Position	p-value	Gene name	Gene position
6	31311271	7.88×10^{-7}	<i>BTC</i>	31225858–31278471
9	46247284	3.19×10^{-13}	<i>LRPPRC</i>	46152562–46258899
9	97711609	1.29×10^{-13}	<i>OTOF</i>	97665582–97696329
9	97711609	1.29×10^{-13}	<i>DRC1</i>	97697741–97740010
9	97711609	1.29×10^{-13}	<i>SELENOI</i>	97753607–97797598
11	58069337	1.86×10^{-10}	<i>OTX2</i>	57983921–57991821
13	80629523	6.56×10^{-15}	<i>CNTN5</i>	79939901–80974967
17	28641668	6.87×10^{-13}	<i>TMC3</i>	28497725–28556487
21	57816348	4.02×10^{-12}	<i>PMAIP1</i>	57811222–57813908
24	3313732	3.63×10^{-8}	<i>H2BC6</i>	3223813–3224193
24	3313732	3.36×10^{-8}	<i>H2AC8</i>	3248645–3249037
24	3313732	3.36×10^{-8}	<i>H4C8</i>	3319662–3319973
24	3313732	3.36×10^{-8}	<i>BTN2A2</i>	3411192–3422262
24	6221834	7.02×10^{-11}	<i>POLR1H</i>	6226933–6230864
24	6221834	7.02×10^{-11}	<i>PPP1R11</i>	6235327–6237209
24	6221834	7.02×10^{-11}	<i>MOG</i>	6188533–6198769
24	6221834	7.02×10^{-11}	<i>RNF39</i>	6241382–6246935
24	6221834	7.02×10^{-11}	<i>TRIM40</i>	6318264–6330272
24	6221834	7.02×10^{-11}	<i>GABBR1</i>	6134269–6164856

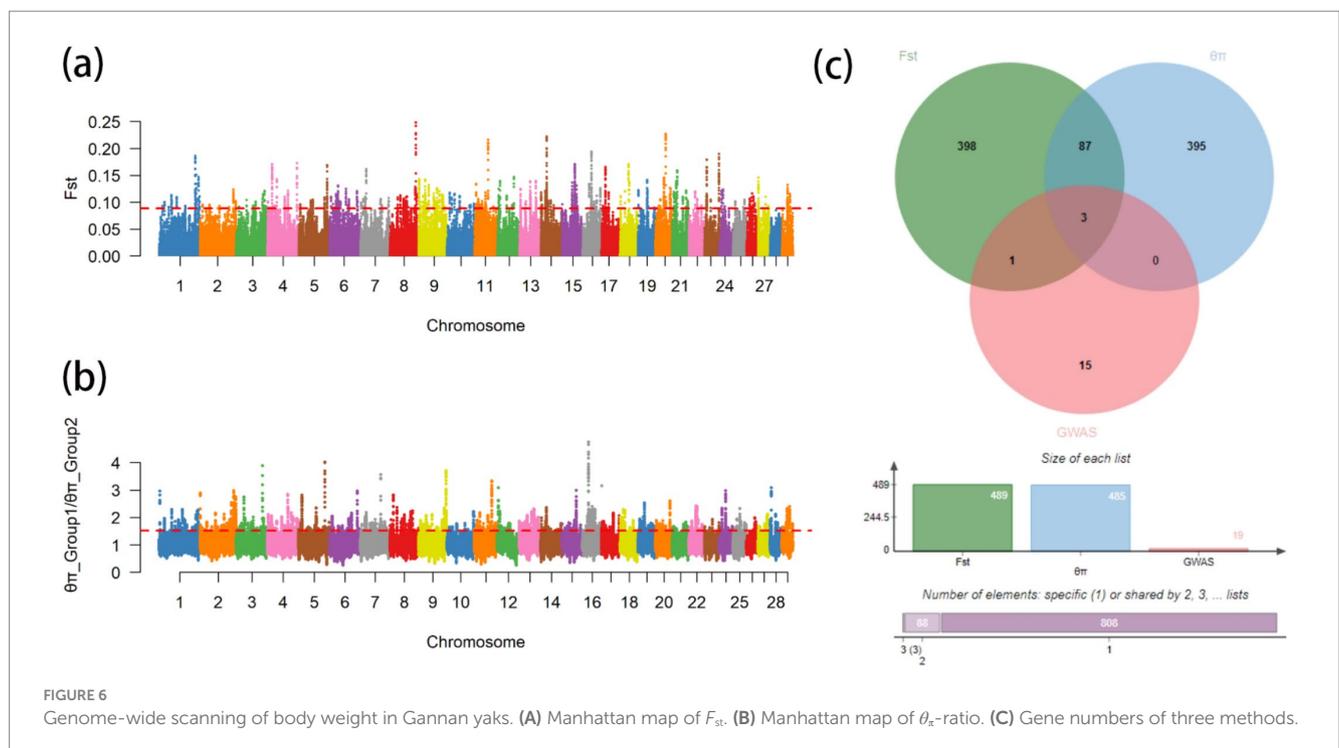
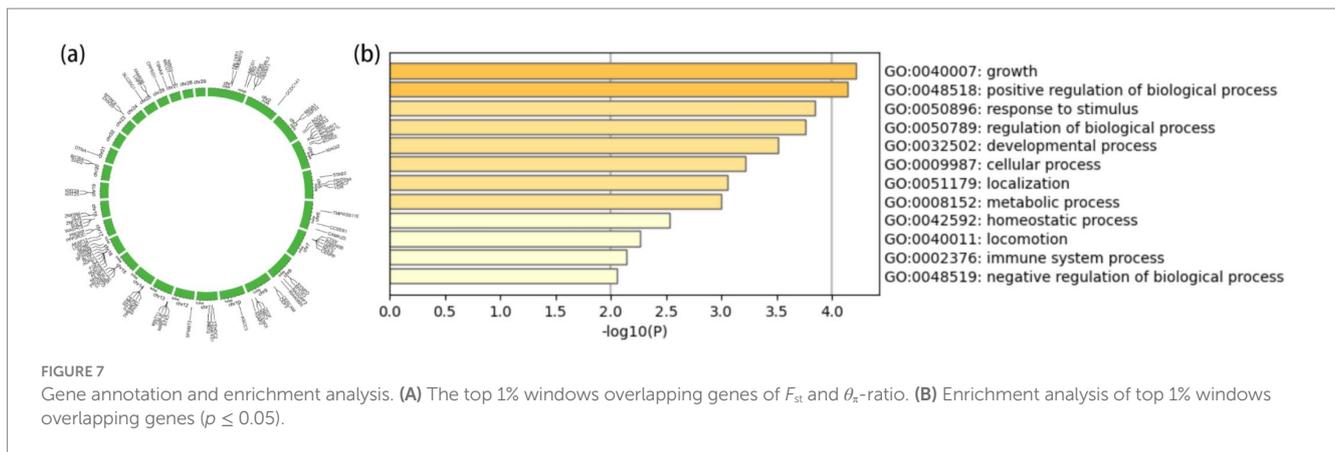


FIGURE 6 Genome-wide scanning of body weight in Gannan yaks. (A) Manhattan map of F_{st} . (B) Manhattan map of θ_{π} -ratio. (C) Gene numbers of three methods.

that the mutation of *DRC1* (T genotype of *Chr9* 97711609) and the mutation of *SELENOI* (C genotype of *Chr9* 97711609) may be associated with BW. Finally, all haplotypes of 247 Gannan yaks at these two sites were extracted, and the difference testing was

carried out according to different haplotype groups combined with BW data (Figures 8E,F). The results showed that the mean BW of homozygous mutant-type (TT of *DRC1* and CC of *SELENOI*) was the largest. The mean BW of homozygous wild-type (GG of *DRC1*



and TT of *SELENOI*) was the smallest. The difference in BW between the two groups was significant (Figures 8F,G), reinforcing the association of *DRC1* and *SELENOI* mutations with BW in Gannan yaks.

4 Discussion

The genetic diversity and BW traits of Gannan yaks hold significant implications for their breeding and conservation. Historically, Gannan yaks have been managed under mixed grazing systems in natural grasslands with limited selective breeding efforts. Unlike modern livestock species such as beef cattle, where systematic breeding programs have achieved substantial genetic improvement, yak breeding has faced challenges due to a lack of pedigree records, reliance on traditional management practices, and limited technological intervention. This study's findings offer valuable insights into the genetic underpinnings of BW in Gannan yaks and provide a molecular basis for advancing their breeding programs.

ROH lengths offer insight into inbreeding history: shorter ROHs indicate older inbreeding, while longer ROHs point to more recent, closer inbreeding events (18). This study revealed that Gannan yaks possess higher genetic diversity compared to other yak breeds, as indicated by the shortest runs of homozygosity (ROH) and the fastest LD decay rates. These findings suggest that Gannan yaks have experienced less recent inbreeding and maintain a relatively large effective population size. Within the Gannan yak population, the high BW group (Group 1) demonstrated even greater genetic diversity than the low BW group (Group 2). This variability highlights the potential of utilizing existing genetic resources within the breed to improve growth traits through selective breeding. The identification of high genetic diversity in Gannan yaks aligns with their historical reliance on natural grazing, which has likely preserved genetic variation due to the absence of intensive artificial selection. However, this diversity also reflects the underdeveloped state of yak breeding programs. Systematic efforts to harness this genetic diversity through marker-assisted selection and GWAS could enhance the productivity of Gannan yaks while retaining their adaptability to the harsh plateau environment.

The study identified 19 genes associated with BW through GWAS, with notable candidates including *LRPPRC*, *PMAIP1*, *PPP1R11*, and *GABBR1*. These genes are involved in critical biological processes such as muscle development, fat metabolism, and cell proliferation. *LRPPRC* plays a role in regulating cytoskeletal network dynamics (19). It influences conformation traits in Danish pig breeds by modulating bone and muscle development (20). *PMAIP1* has been linked to human obesity (21). *PPP1R11* encodes phosphatase 1 regulatory (inhibitor) subunit 11, it may be involved in Hippo signaling in apoptosis and cell proliferation and has been linked to *VLDL* particle concentration in plasma (22, 23). *GABBR1*, which encodes a receptor for gamma-aminobutyric acid (GABA) involved in the cAMP signaling pathway, plays a role in fatty acid β -oxidation and degradation. This gene has also been linked to fat depot-specific adipose tissue macrophage infiltration in obesity (24). These findings suggest that the genetic architecture of BW in Gannan yaks is influenced by pathways that overlap with those in other livestock species, providing an opportunity to leverage cross-species knowledge for yak breeding.

The genome-wide scanning analysis further identified 90 candidate genes, including *FGF2*, *SHH* and *WNT11*, which are enriched in growth and developmental processes. *SHH* (*Sonic Hedgehog*) promotes the growth of intestinal epithelium by activating the SHH-WNT signaling axis (25). During the morphological changes occurring in osteoblast differentiation, the epigenetic machinery in modulating the *BMP7*-induced osteogenic phenotype by influencing the activity of *Shh-related* genes (26). Shh-induced Akt phosphorylation promotes muscle cell proliferation and differentiation, and *SHH* is critical for controlling specific ventral muscle formation (27, 28). Fibroblast growth factor 2 (*FGF2*) is a potent mitogenic growth factor expressed in limb bud mesenchyme, chondrocytes, osteoblasts, osteoclasts, and adipocytes, promoting osteoblast differentiation by regulating *BMP2* and *ATF4* as well as Wnt/ β -catenin pathway (29, 30). *FGF2-null* mice exhibit osteopenia and reduced bone remodeling (31–33). In addition, N-acetylated *HS* mimics promote *FGF2* signaling, thereby promoting muscle stem cell expansion (34). *WNT11* acts as a directional cue to organize myotube elongation in the myotome via the noncanonical Wnt/planar cell polarity pathway (35), co-expression of *Akt1* and *Wnt11* significantly boosts the proliferation and growth of mesenchymal stem cells, while

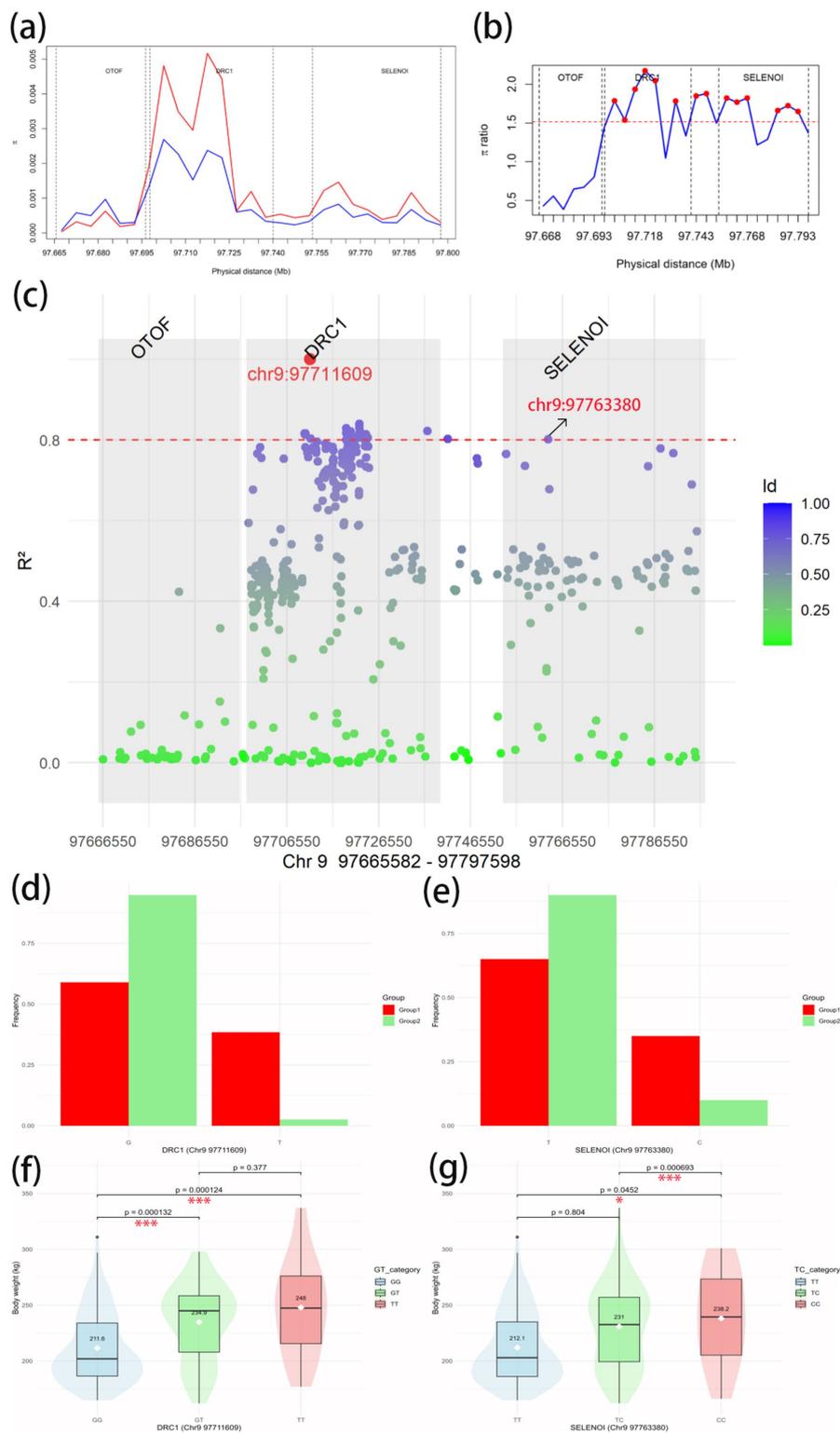


FIGURE 8

Candidate gene region analysis. **(A)** Nucleotide diversity in *Chr9* 97665582–97797598 regions. Red represents the Group 1; blue represents the Group 2. **(B)** Nucleotide diversity ratio (Group 1/Group 2) in *Chr9* 97665582–97797598. The red dashed line is the threshold line for the top 1% window of the whole-genome scan θ_{π} -ratio. Red dots indicate windows above the threshold line. The vertical dotted lines in the figure are the dividing lines between genes. **(C)** Linkage imbalance with significant GWAS site *Chr9* 97711609 to other SNPs in *Chr9* 97665582–97797598 regions of 247 Gannan yaks. The dotted red line is the strong linkage threshold line of $R^2 = 0.8$. The gray box shows the range of each of the three genes. **(D)** The allele frequencies of G and T at *Chr9* 97711603 site in Group 1 and Group 2. G is the wild-type and T is the mutant-type. **(E)** The allele frequencies of T and C at *Chr9* 97763380 site in Group 1 and Group 2. T is the wild-type and C is the mutant-type. **(F)** Body weight phenotypic differences between haplotypes of *DRC1* in *Chr9* 97711603 site. **(G)** Body weight phenotypic differences between haplotypes of *SELENOI* in *Chr9* 97763380 site. *Significant at $p < 0.05$, **significant at $p < 0.01$, and ***significant at $p < 0.001$.

Wnt11 overexpression synergistically promotes chondrogenic differentiation with TGF- β (36, 37). These genes underscore the complex genetic regulation of BW and highlight targets for future functional studies.

Through the analysis of the three overlapping genes, *DRC1* and *SELENOI* were identified as key candidate genes for BW in Gannan yak. Other studies demonstrated that *DRC1* is a critical gene in cell cycle-regulated essential for DNA replication and necessary for the S-phase checkpoint, including the proper activation of Rad53 kinase in response to DNA damage and replication interruptions (38). Furthermore, *DRC1* expression in follicular dendritic cells (FDCs) underscores its association with muscle cells. FDCs expressing *DRC1* also express α -smooth muscle actin (α -SM actin) and form contractile stress fibers, key features of muscle cells, particularly myofibroblasts (39). *SELENOI* plays a role in maintaining redox-related calcium homeostasis and helps mitigate muscle degeneration and related kyphosis (40–42). Its expression is also upregulated in the skeletal muscles of chickens in response to selenium deficiency in the diet (43). These findings indicated that *DRC1* and *SELENOI* play roles in DNA replication, muscle cell function, and metabolic homeostasis, making them promising targets for improving BW in Gannan yaks.

The identification of candidate genes for BW provides a scientific foundation for addressing long-standing challenges in Gannan yak breeding. Unlike beef cattle, where breeding programs have been optimized for traits such as growth rate and feed efficiency, Gannan yak breeding has lagged due to limited genomic resources and infrastructure. This study bridges this gap by identifying molecular markers that can be integrated into selective breeding programs. Furthermore, the high genetic diversity observed in Gannan yaks suggests that *in situ* conservation strategies can be complemented by genetic improvement efforts. By prioritizing the identified candidate genes in breeding programs, it is possible to enhance BW traits without compromising the genetic diversity that underpins the breed's adaptability. The findings also provide a framework for developing genomic selection models tailored to the unique characteristics of Gannan yaks, accelerating the pace of genetic improvement.

This study demonstrates the utility of combining GWAS, genome-wide scanning, functional annotation and SNP association analyses to dissect complex traits in livestock species with limited breeding histories. The findings highlight the potential of genomic tools to revolutionize yak breeding, moving beyond traditional practices to data-driven approaches. However, realizing these benefits will require sustained investment in genomic infrastructure, farmer education, and policy support. Future research should focus on validating the functional roles of the identified genes and exploring gene-environment interactions that may influence BW. Additionally, integrating genomic data with phenotypic records and environmental variables could provide a holistic understanding of the factors shaping BW traits in Gannan yaks. These efforts will be essential for ensuring the sustainability and competitiveness of yak husbandry in the face of changing environmental and market conditions.

In conclusion, this study lays the groundwork for transforming Gannan yak breeding through genomic insights. By leveraging the genetic diversity and candidate genes identified here, it is possible to enhance the economic and cultural value of this iconic breed while preserving its ecological resilience.

Data availability statement

The original contributions presented in the study are publicly available. These data can be found in <https://doi.org/10.6084/m9.figshare.29143589>.

Ethics statement

The animal studies were approved by the regulations of Lanzhou Institute of Husbandry and Pharmaceutical Sciences of the Chinese Academy of Agricultural Sciences. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

MZ: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. CH: Methodology, Writing – original draft, Writing – review & editing. QB: Formal analysis, Methodology, Writing – review & editing. TW: Formal analysis, Writing – review & editing. XM: Writing – review & editing. GM: Methodology, Writing – review & editing. XZ: Methodology, Writing – review & editing. QZ: Methodology, Writing – review & editing. MC: Conceptualization, Writing – review & editing. CL: Conceptualization, Writing – review & editing. XG: Conceptualization, Writing – review & editing. PB: Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – review & editing. PY: Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was jointly supported by the Major Science and Technology Projects in Gansu Province (Grant Nos. 21ZD10NA001 and GZGG-2021-1) Breeding promotion and germplasm resources exploration of Gannan yak.

Conflict of interest

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Yan P, Liang C. Chinese yak. 1st ed. Beijing: China Agricultural Science and Technology Press (2019). 591 p.
- Wang J, Li X, Peng W, Zhong J, Jiang M. Genome-wide association study of body weight trait in yaks. *Animals*. (2022) 12:1855. doi: 10.3390/ani12141855
- Xu C, Wu S, Zhao W, Mipam T, Liu J, Liu W, et al. Differentially expressed microRNAs between cattleyak and yak testis. *Sci Rep*. (2018) 8:592. doi: 10.1038/s41598-017-18607-0
- Gomez Proto G, Mancin E, Sartori C, Mantovani R. Unraveling inbreeding patterns and selection signals in Alpine Grey cattle. *Animal*. (2024) 18:101159. doi: 10.1016/j.animal.2024.101159
- Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, et al. The yak genome and adaptation to life at high altitude. *Nat Genet*. (2012) 44:946–9. doi: 10.1038/ng.2343
- Ji Q, Xin J, Chai Z, Zhang C, Dawa Y, Luo S, et al. A chromosome-scale reference genome and genome-wide genetic variations elucidate adaptation in yak. *Mol Ecol Resour*. (2021) 21:201–11. doi: 10.1111/1755-0998.13236
- Dai R, Huang C, Wu X, Ma X, Chu M, Bao P, et al. Copy number variation (CNV) of the *AHR* gene in the Ashidan yak and its association with growth traits. *Gene*. (2022) 826:146454. doi: 10.1016/j.gene.2022.146454
- Huang C, Ge F, Ren W, Zhang Y, Wu X, Zhang Q, et al. Copy number variation of the *HPGDS* gene in the Ashidan yak and its associations with growth traits. *Gene*. (2021) 772:145382. doi: 10.1016/j.gene.2020.145382
- Ren W, Huang C, Ma X, La Y, Chu M, Guo X, et al. Association of *HSF1* gene copy number variation with growth traits in the Ashidan yak. *Gene*. (2022) 842:146798. doi: 10.1016/j.gene.2022.146798
- Zhang Z, Chu M, Bao Q, Bao P, Guo X, Liang C, et al. Two different copy number variations of the *SOX5* and *SOX8* genes in yak and their association with growth traits. *Animals*. (2022) 12:1587. doi: 10.3390/ani12121587
- Li X, Huang C, Liu M, Dai R, Wu X, Ma X, et al. Copy number variation of the *SOX6* gene and its associations with growth traits in Ashidan yak. *Animals*. (2022) 12:3074. doi: 10.3390/ani12223074
- Aldana R, Freed D. Data processing and germline variant calling with the sentieon pipeline. In: C Ng and S Piscooglio, editors. Variant calling: methods and protocols. New York, NY: Springer (2022). 1–19.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. (2007) 81:559–75. doi: 10.1086/519795
- Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. (2019) 35:1786–8. doi: 10.1093/bioinformatics/bty875
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics*. (2021) 19:619–28. doi: 10.1016/j.gpb.2020.10.007
- Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics*. (2014) 47:11.12.1–11.12.34. doi: 10.1002/0471250953.bi1112s47
- Han R, Han L, Zhao X, Wang Q, Xia Y, Li H. Haplotype-resolved genome of sika deer reveals allele-specific gene expression and chromosome evolution. *Genomics Proteomics Bioinformatics*. (2023) 21:470–82. doi: 10.1016/j.gpb.2022.11.001
- Zhao Z, Sun Y, Tang J, Yang Y, Xu X. LRPPRC regulates malignant behaviors, protects mitochondrial homeostasis, mitochondrial function in osteosarcoma and derived cancer stem-like cells. *BMC Cancer*. (2023) 23:1–11. doi: 10.1186/s12885-023-11443-8
- Le TH, Christensen OF, Nielsen B, Sahana G. Genome-wide association study for conformation traits in three Danish pig breeds. *Genet Sel Evol*. (2017) 49:12. doi: 10.1186/s12711-017-0289-2
- Voisin S, Almén MS, Zheleznyakova GY, Lundberg L, Zarei S, Castillo S, et al. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome Med*. (2015) 7:103. doi: 10.1186/s13073-015-0225-4
- Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyytikäinen L-P, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet*. (2012) 44:269–76. doi: 10.1038/ng.1073
- Piórkowska K, Żukowski K, Ropka-Molik K, Tyra M, Gurgul A. A comprehensive transcriptome analysis of skeletal muscles in two polish pig breeds differing in fat and meat quality traits. *Genet Mol Biol*. (2018) 41:125–36. doi: 10.1590/1678-4685-GMB-2016-0101
- Hwang I, Jo K, Shin KC, Kim JI, Ji Y, Park YJ, et al. GABA-stimulated adipose-derived stem cells suppress subcutaneous adipose inflammation in obesity. *Proc Natl Acad Sci USA*. (2019) 116:11936–45. doi: 10.1073/pnas.1822067116
- Orzechowska-Licari EJ, Bialkowska AB, Yang VW. Sonic hedgehog and WNT signaling regulate a positive feedback loop between intestinal epithelial and stromal cells to promote epithelial regeneration. *Cell Mol Gastroenterol Hepatol*. (2023) 16:607–42. doi: 10.1016/j.jcmgh.2023.07.004
- GDS F, Alves Dos Santos EA, de Camargo Andrade AF, Zambuzzi WF, da Silva RAF. BMP7-induced osteoblast differentiation requires hedgehog signaling and involves nuclear mechanisms of gene expression control. *Cell Biol Int*. (2024) 48:939–50. doi: 10.1002/cbin.12161
- Elia D, Madhala D, Ardon E, Reshef R, Halevy O. Sonic hedgehog promotes proliferation and differentiation of adult muscle cells: involvement of MAPK/ERK and PI3K/Akt pathways. *Biochim Biophys Acta*. (2007) 1773:1438–46. doi: 10.1016/j.bbamcr.2007.06.006
- Anderson C, Williams VC, Moyon B, Daubas P, Tajbakhsh S, Buckingham ME, et al. Sonic hedgehog acts cell-autonomously on muscle precursor cells to generate limb muscle diversity. *Genes Dev*. (2012) 26:2103–17. doi: 10.1101/gad.187807.112
- Fei Y, Xiao L, Doetschman T, Coffin DJ, Hurley MM. Fibroblast growth factor 2 stimulation of osteoblast differentiation and bone formation is mediated by modulation of the Wnt signaling pathway*. *J Biol Chem*. (2011) 286:40575–83. doi: 10.1074/jbc.M111.274910
- Fei Y, Xiao L, Hurley MM. The impaired bone anabolic effect of PTH in the absence of endogenous FGF2 is partially due to reduced ATF4 expression. *Biochem Biophys Res Commun*. (2011) 412:160–4. doi: 10.1016/j.bbrc.2011.07.066
- Su N, Jin M, Chen L. Role of FGF/FGFR signaling in skeletal development and homeostasis: learning from mouse models. *Bone Res*. (2014) 2:14003. doi: 10.1038/boneres.2014.3
- Montero A, Okada Y, Tomita M, Ito M, Tsurukami H, Nakamura T, et al. Disruption of the fibroblast growth factor-2 gene results in decreased bone mass and bone formation. *J Clin Invest*. (2000) 105:1085–93. doi: 10.1172/JCI8641
- Homer-Bouthiette C, Doetschman T, Xiao L, Hurley MM. Knockout of nuclear high molecular weight FGF2 isoforms in mice modulates bone and phosphate homeostasis*. *J Biol Chem*. (2014) 289:36303–14. doi: 10.1074/jbc.M114.619569
- Ghadiali RS, Guimond SE, Turnbull JE, Pisconti A. Dynamic changes in heparan sulfate during muscle differentiation and ageing regulate myoblast cell fate and FGF2 signalling. *Matrix Biol*. (2017) 59:54–68. doi: 10.1016/j.matbio.2016.07.007
- Gros J, Serrallbo O, Marcelle C. WNT11 acts as a directional cue to organize the elongation of early muscle fibres. *Nature*. (2009) 457:589–93. doi: 10.1038/nature07564
- Chen B, Chen X, Liu C, Li J, Liu F, Huang Y. Co-expression of Akt1 and Wnt11 promotes the proliferation and cardiac differentiation of mesenchymal stem cells and attenuates hypoxia/reoxygenation-induced cardiomyocyte apoptosis. *Biomed Pharmacother*. (2018) 108:508–14. doi: 10.1016/j.biopha.2018.09.047
- Liu S, Zhang E, Yang M, Lu L. Overexpression of Wnt11 promotes chondrogenic differentiation of bone marrow-derived mesenchymal stem cells in synergism with TGF- β . *Mol Cell Biochem*. (2014) 390:123–31. doi: 10.1007/s11010-014-1963-0
- Wang H, Elledge SJ. DRC1, DNA replication and checkpoint protein 1, functions with DPB11 to control DNA replication and the S-phase checkpoint in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*. (1999) 96:3824–9. doi: 10.1073/pnas.96.7.3824
- Muñoz-Fernández R, Blanco FJ, Frecha C, Martín F, Kimatrai M, Abadía-Molina AC, et al. Follicular dendritic cells are related to bone marrow stromal cell progenitors and to myofibroblasts. *J Immunol*. (2006) 177:280–9. doi: 10.4049/jimmunol.177.1.280
- Arbogast S, Ferreiro A. Selenoproteins and protection against oxidative stress: selenoprotein N as a novel player at the crossroads of redox signaling and calcium homeostasis. *Antioxid Redox Signal*. (2010) 12:893–904. doi: 10.1089/ars.2009.2890

41. Castets P, Bertrand AT, Beuvin M, Ferry A, Le Grand F, Castets M, et al. Satellite cell loss and impaired muscle regeneration in selenoprotein N deficiency. *Hum Mol Genet.* (2011) 20:694–704. doi: 10.1093/hmg/ddq515
42. Rederstorff M, Castets P, Arbogast S, Lainé J, Vassilopoulos S, Beuvin M, et al. Increased muscle stress-sensitivity induced by selenoprotein N inactivation in mouse: a mammalian model for SEPNI-related myopathy. *PLoS One.* (2011) 6:e23094. doi: 10.1371/journal.pone.0023094
43. Zhang J, Zhang Z-W, Shan A-S, Xu S. Effects of dietary selenium deficiency or excess on gene expression of selenoprotein N in chicken muscle tissues. *Biol Trace Elem Res.* (2014) 157:234–41. doi: 10.1007/s12011-014-9893-y