

RESEARCH

Open Access



# Comparative whole-genome resequencing to uncover selection signatures linked to litter size in Hu Sheep and five other breeds

Tao Zhong<sup>1\*†</sup>, Dunying Hou<sup>1†</sup>, Qianjun Zhao<sup>2</sup>, Siyuan Zhan<sup>1</sup>, Linjie Wang<sup>1</sup>, Li Li<sup>1</sup>, Hongping Zhang<sup>1</sup>, Wei Zhao<sup>3</sup>, Shizhong Yang<sup>4</sup> and Lili Niu<sup>1</sup>

## Abstract

Hu sheep (HS), a breed of sheep carrying the *FecB* mutation gene, is known for its “year-round estrus and multiple births” and is an ideal model for studying the high fecundity mechanisms of livestock. Through analyzing and comparing the genomic selection features of Hu sheep and other sheep breeds, we identified a series of candidate genes that may play a role in Hu sheep’s high fecundity mechanisms. In this study, we conducted whole-genome resequencing on six breeds and screened key mutations significantly correlated with high reproductive traits in sheep. Notably, the *CC2D1B* gene was selected by the fixation index ( $F_{ST}$ ) and the cross-population composite likelihood ratio (XP-CLR) methods in HS and other five breeds. It was worth noting that the *CC2D1B* gene in HS was different from that in other sheep breeds, and seven missense mutations have been identified. Furthermore, the linkage disequilibrium (LD) analysis revealed a strong linkage disequilibrium in this specific gene region. Subsequently, by performing different grouping based on *FecB* genotypes in Hu sheep, genome-wide selective signal analysis screened several genes related to reproduction, such as *BMPR1B* and *PPM1K*. Besides,  $F_{ST}$  analysis identified functional genes related to reproductive traits, including *RHEB*, *HSPA2*, *PPP1CC*, *HVCN1*, and *CCDC63*. Additionally, a missense mutation was found in the *CCDC63* gene and the haplotype was different between the high reproduction (HR) group and low reproduction (LR) group in HS. In summary, we discovered genetic differentiation among six distinct breeding sheep breeds at the whole genome level. Additionally, we identified a set of genes which were associated with reproductive performance in Hu sheep and visualized how these genes differed in different breeds. These findings laid a theoretical foundation for understanding genetic mechanisms behind high prolific traits in sheep.

**Keywords** Hu sheep, Whole-genome resequencing, Selection signatures, Litter size

<sup>†</sup>Tao Zhong and Dunying Hou contributed equally to this work.

\*Correspondence:

Tao Zhong  
zhongtao@sicau.edu.cn

<sup>1</sup>Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu 611130, China

<sup>2</sup>Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China

<sup>3</sup>College of Animal Science, Xichang University, Xichang 615013, China

<sup>4</sup>Academy of Agricultural Sciences Liangshan, Xichang 615000, China



## Introduction

Sheep (*Ovis aries*) have been a foundational source of meat, wool, and milk for humans, representing one of the most economically significant livestock for thousands of years [1]. The reproductive trait holds significant economic value for sheep production, regulated by heredity [2], hormone [3], environment [4], and managing factors. Previous studies have demonstrated the pivotal role of genetic factors in determining the reproductive performances of sheep. Elucidating the genetic mechanisms underlying high prolificacy is crucial for improving sheep production through the utilization of prolific resources. The litter size in sheep was influenced by a combination of major genes and polygenes. Currently, multiple major genes affecting prolificacy have been reported in sheep [5]. Previous studies have demonstrated a positive correlation between litter size and the genotypes of *FecB* [6]. The distribution of *FecB* genotypes was various among breeds and strains [7, 8]. In the high-fertility Hu sheep, some ewes harboring the *FecB* homozygous mutation (c.746 A>G) that delivered a single lamb, indicating additional genetic factors resulted in a reduction in litter sizes. Screening for fertility genes is helpful for marker-assisted breeding to increase the performance of reproduction in sheep. Uncovering the genetic mechanism of sheep prolificacy is conducive to enhancing the reproduction of Chinese sheep breeds.

The genetic mechanisms of morphological and agronomic traits in sheep have been gradually revealed and continue to be studied with the completion of the sheep reference genome [9]. Studies have focused on various aspects, including the domestication and chromosome evolution in wild sheep [10]. Additionally, researchers have delved into important traits, such as reproduction, coat color [11], horn phenotype [12], tail fat deposition [13], and body size [14]. Whole-genome resequencing could identify the selective traits behind phenotypic differences and reveal the genetic basis of complex traits [13]. Identifying the genetic variations responding to different phenotypes could clarify the genetic mechanisms underlying high productivity in varieties. The acquisition of genome-wide genetic mutations provided a solid foundation for establishing genomic selective breeding [15]. The primary work is to further explore reproduction-related genes and functional loci to supply favorable conditions for exploring the genetic underpinnings of high fecundity and new variety breeding. While the main relevant major genes affecting the reproductive traits have been identified, such as *BMP15* [16, 17], *GDF9* [17, 18], *B4GALNT2* [5], and *KISS1R* [19], their genetic roles in most sheep breeds remained unclear. In addition, the candidate genes affecting litter size varied among breeds and resulted from the combined effects of multiple candidate genes. Thus, further investigation is necessary to

uncover the key genes regulating reproductive traits in sheep [20–22]. Hu sheep, a prolific native breed in China, serve as valuable animal models for understanding the genetic underpinnings of prolific sheep breeds [23]. Furthermore, analyzing the genome data in Hu sheep could provide new insight into differences in reproduction performance, potentially enabling the replication of the multiple birth characteristics of Hu sheep in other sheep breeds [24].

Lacking evidence to clarify the regulatory mechanisms and candidate genes related to reproductive traits, we performed the whole-genome resequencing of Hu sheep and Liangshan Black sheep (LB), and jointly analyzed with the genome-wide data of Bamei Mutton sheep (BM), White Xizang sheep (WX), Oula sheep (OL), and Poll Dorset sheep (PD) aimed to explore the molecular basis of the high fertility trait, as well as conducting a comparative analysis of genome-wide DNA SNPs among individuals with different *FecB* genotypes. We identified genomic regions and key mutations significantly associated with high fertility traits in Hu sheep and validated in various breeds with different fertility levels. In addition, our findings contribute to the exploration of high-quality genes related to high fertility in Hu sheep and provide insight into the molecular regulatory mechanisms underlying reproduction among different sheep breeds.

## Materials and methods

### Sampling collection and DNA extraction

274 ewes of Hu sheep were collected from the sheep farm of the Yuexi Oriental Agricultural Development Co., LTD (Liangshan, Sichuan, China). 53 female Liangshan Black sheep were randomly selected from the Butuo County Black Sheep Breeding Farm (Liangshan, Sichuan, China). Whole blood was collected from the jugular vein and kept in EDTA vacutainer tubes. The genomic DNA of blood sample was extracted using a Genomic DNA Kit (TansGen Biotech, Beijing, China). The integrity and concentration of DNA samples were assessed using 1% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Extracted DNA samples were stored at -20 °C prior to further analysis.

### Genotyping of the c.746 A>G locus of *FecB* gene

A fragment of 328 bp harbored the c.746 A>G locus was amplified by the primer pair (Forward 5'-CAGATGGT GAAACAGATTG-3', Reverse 5'-CAAGTCCACCATC CATTC-3'), which was designed by Primer Premier 5.0 software and synthesized by Sangon Biotech Co., Ltd (Shanghai, China). Polymerase chain reaction (PCR) was conducted on a BIO-RAD C1000Touch™ Thermal Cycler (CA, USA) under the following conditions: 95 °C for 3 min and 34 cycles of 30 s at 95 °C, 30 s at 55 °C, 23 s

at 72 °C, and a final step of 5 min at 72 °C, then stored at 4 °C. The PCR products were purified and sequenced by an ABI 3730 Sequencer (Thermo Fisher Scientific, Waltham, MA, USA) to determine the *FecB* genotype (c.746 A>G) in all HS and LB individuals. In addition, the genotype of the c.746 A>G locus in the other joined analyzed breeds (BM, WX, OL, and PD) were determined by the whole genome resequencing data.

#### Whole genome resequencing and data processing

29 HS and 19 LB were randomly selected for whole genome resequencing. The paired-end sequencing libraries with an insert size of ~350 bp fragments for each individual were constructed to ~10× raw coverage using the BGISEQ DNBSEQ-T7 platform (BGI Inc., Shenzhen, China) according to the manufacturer's protocol at Novogene Bioinformatics Technology Co., Ltd (Beijing, China). In addition, resequencing data of BM, WX, OL, and PD were obtained from the previous studies [25, 26].

All fastq files were processed through the use of fastp (v0.19.5) to obtain clean reads. The following filter criteria were employed to eliminate adapters and low-quality bases: reads with more than 10% unknown nucleotides (N), reads with over 50% low-quality bases (Q-value<5), and reads with more than 10 nucleotides aligned to the adaptor sequence with a maximum of two mismatches. The resulting clean reads were then mapped to the sheep Oar\_v4.0 reference genome utilizing Burrows-Wheeler Aligner v0.7.17 (BWA) algorithm with default parameters [27]. Mapping files were converted to BAM files and sorted using SAMtools (v1.9). Single nucleotide polymorphism (SNP) calling was performed using the Bayesian method in the GATK package (v4.1.3.0–0) with the subsequent filtering criteria: QD<10.0, ReadPos RankSum < -8.0, FS>10, QUAL<30, and DP<4. VCFtools (v0.1.14) was utilized to generate a raw VCF file, which was then filtered to create a high-quality VCF file with retained SNPs for further analysis. The filtering criteria for the high-quality VCF file were as follows: (1) %QUAL<100 and (2) INFO/DP<5. To annotate the SNPs and indels, ANNOVAR was employed, utilizing gene models from GFF annotation [28].

#### Genetic diversity and runs of homozygosity (ROH)

Nucleotide diversity ( $\pi$ ) for each breed was investigated using VCFtools v0.1.16 with a 40 kb non-overlapping window across all autosomes [29]. PLINK (--hardy) was used to estimate observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ).  $H_e$  and  $H_o$  estimates for individuals in each breed were averaged across all SNPs. Long homozygous fragments were scanned by using PLINK according to the following parameters: --homozyg-window-snp 50 --homozyg-snp 50 --homozyg-kb 300 --homozyg-density 50 --homozyg-gap 1000

--homozyg-window-missing 5 --homozyg-window-threshold 0.05 --homozyg-window-het 1 [30]. The genomic inbreeding coefficient based on ROH ( $F_{ROH}$ ) is the ratio of the total length of ROH fragments in the genome to the total length of the permanent genome, and each ROH is classified according to its physical length as follows: 0–1, 1–4, and  $\geq 4$  Mb.

#### Population structure analysis

Population genomics analyses was used to explore the genetic relationship of individuals from HS, LB, BM, WX, OL, and PD, SNPs that did not meet one of the following criteria (--maf 0.05 --max-missing 0.9) were filtered by PLINK v1.9 [31]. Additionally, the linkage disequilibrium (LD) of each sheep population was removed with the criteria: --indep-pairwise 50 10 0.1. Principal components analysis (PCA) was applied to visualize patterns in relationships between six sheep breeds. Furthermore, population structure was examined using ADMIXTURE v.1.23 to estimate the cross-error for genetic clustering with the ancestral clusters ( $K$ ) ranging from 2 to 5 and each  $K$  was run the analysis for 100 times [32]. The SNPs were used to calculate distance matrix by the software VCF-2Dis to construct the phylogenetic tree. The topological structure was displayed by the Interactive Tree of Life (ITOL) tool (<https://itol.embl.de/>) [33]. The linkage disequilibrium (LD)  $r^2$  with physical distance between SNPs for each breed was calculated using PopLDdecay with default parameters [34].

#### Selective signatures related to litter size

We identified genome-wide selective sweeps during all breeds based on  $F_{ST}$  analysis using VCFtools. The  $F_{ST}$  values were calculated using a sliding window approach, with 80 kb windows and 40 kb sliding steps according to the previous study [35]. Windows with  $F_{ST}$  values which were >0.25 were defined as the putatively selected genomic regions. For the fecundity analysis of Hu sheep, we combined the LB, OL, WX, BM, and PD into a group (others) and compared them with HS using two different statistics, including  $F_{ST}$  and XP-CLR with size 80 kb windows and 40 kb step. Overlapping regions that  $F_{ST} > 0.25$  [36] and XP-CLR score>2.5 were identified to be candidate regions. The selective sweeps of Hu sheep breeds were also detected by comparison with different litter size. The haplotype blocks based on vcf file were analyzed by LDBlockShow [37]. Moreover, the KOBAS 3.0 (<http://kobas.cbi.pku.edu.cn/>) was utilized to conduct the Kyoto Encyclopaedia of Genes and Genomes (KEGG) analysis.

## Results

### Whole genome resequencing, SNP calling, and genetic diversity

We conducted whole-genome resequencing on a total of 86 sheep samples (Table S1). The sequencing data had an average coverage depth of  $\sim 10.31\times$  for HS and LB,  $\sim 5.76\times$  for the previously characterized 39 individual genomes. After aligning the clean reads to the reference genome, we identified a total of 39,467,233 putative SNPs and 8,677,193 indels in the six breeds, 50.1% of which were located within introns and 41.1% within intergenic regions. A total of 26,664,234 SNPs were retained after quality filtering.

To gain insights into the genetic diversity of the six sheep breeds,  $H_e$ ,  $H_o$ , and the numbers of ROH were estimated based on genotype frequencies (Table 1). The highest  $H_e$  value was observed in HS, indicating a greater level of genetic diversity within the population, while the lowest values were found in WX. The  $H_o$  within the population ranged from 0.1828 to 0.2395, respectively, with the lowest values in BM and the highest in LB. Apart from that, ROH analysis revealed that HS and LB had 14 and 38 long ROH segments, respectively. Meanwhile, shorter ROH segments existed in OL, WX, BM, and PD. In addition, the LD decay pattern and ROH distribution of each group were roughly consistent with the results of nucleotide diversity. (Fig. S1). The distribution of SNPs in each chromosome was visualized in Fig. S2, presenting a non-uniform distribution with wide coverage.

### Population-level analyses of genetic structure and LD decay

Principal component analysis showed that all sheep clustered into four distinct genetic groups that were completely separate from each other, while WX and OL that belonged to Tibetan sheep were clustered together (Fig. 1A). BM and PD were inextricably interwoven as a branch. Population structure analysis, aiming to estimate the proportion of common ancestry among the six breeds, yielded five genetic clusters at the optimal number  $K=5$  (Fig. 1B). We observed that WX and OL were in the same genetic clusters, BM and PD population tended towards completely distinct layers while  $K=4$  and  $K=5$ ,

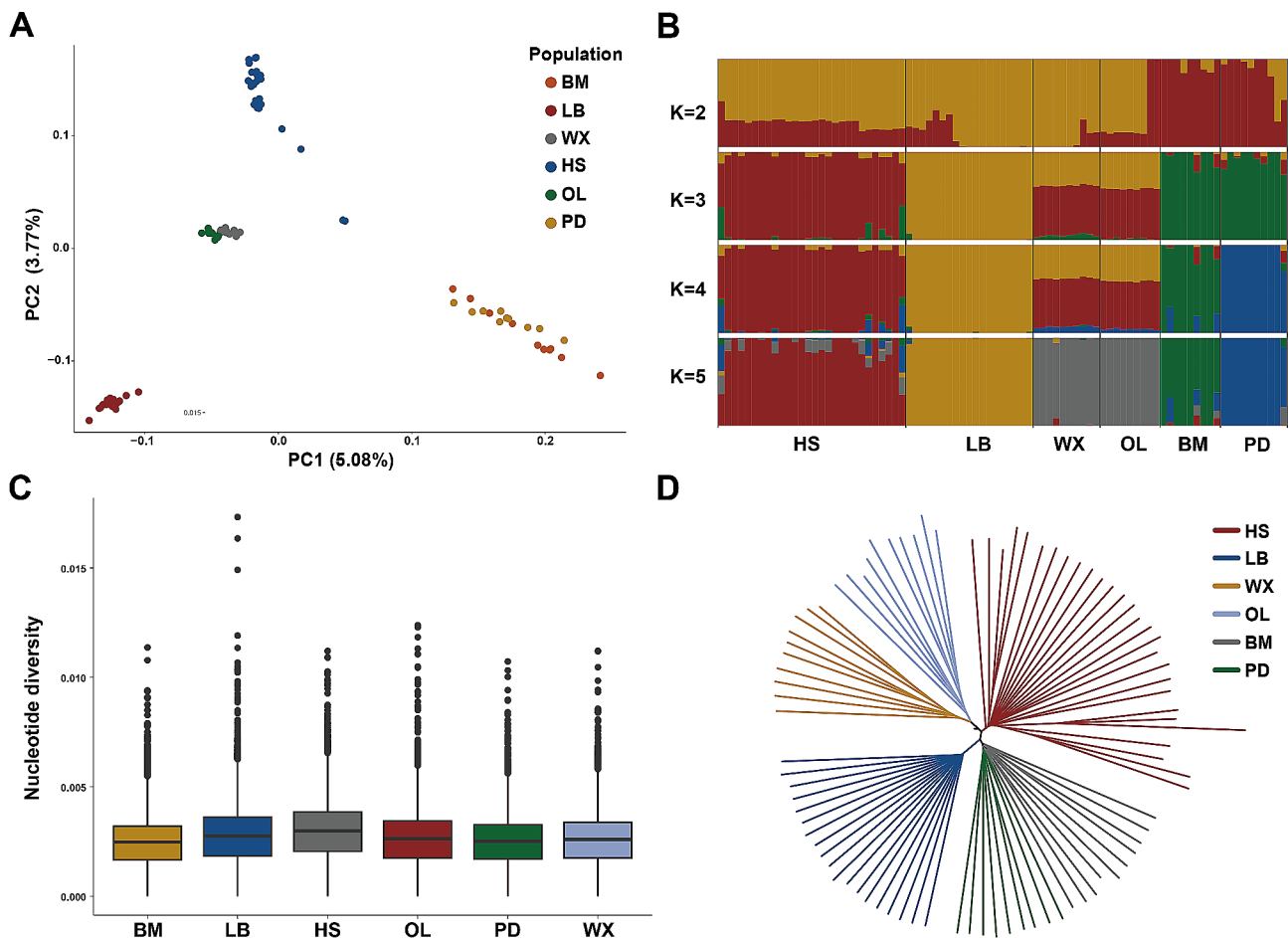
these results were consistent with PCA. To further verify the relationships among breeds, NJ tree results showed that all individuals of the same species formed their own clusters (Fig. 1D). Obviously, HS were separated in a single cluster distinctively, meanwhile, OL and WX were in a branch which formed a main branch with LB. On the other hand, BM and PD were in the same branch, which were in agreement about the result of PCA. LD decay analysis suggested that HS showed minimum LD values at larger physical distances, followed by LB and WX (Fig. S1). The LD decay in OL, PD, and BM had the maximum values. In addition, we observed a relatively low level of nucleotide diversity in BM (0.002448429), the average value of nucleotide diversity in HS (0.002958546) was higher than that in other groups (Fig. 1C).

### Selection signals and candidate genes related to prolificacy

Figure 2A presented the positive genome selections for  $F_{ST}$  in six breeds. Thresholds set at  $F_{ST} > 0.25$  uncovered substantial sites with 71 putative selective sweeps accounting for 211 genes exhibiting positive selection within six breeds, as detailed in Table S2. Furthermore, KEGG Pathway analysis identified the specific pathways in which selected genes were involved (Table S3). We detected candidate genes related to reproduction (e.g., *TSHR*, *THRA*, *CDC25A*, *RARA*), what's more, *MED24* and *THRA* played a role in thyroid hormone signaling pathway which was closely related to reproduction. *RARA* was involved in the estrogen signaling pathway. *CDC25A* was found to be enriched in progesterone-mediated oocyte maturation. *RIPK1* and *TNFAIP3* were found associated with NF-kappa B signaling pathway. Genes associated with disease and immunity, including *GMDS*, *TNFAIP3*, *LRRK2*, and *DDB1*, were identified in several candidate genomic regions under selection. Selective sweep analysis also captured candidate genes associated with adaptation to the plateau, of these, *IL6R* was involved in HIF-1 signaling pathway. Meanwhile, positive selection was detected in *HoxA* and *HoxC* gene cluster. In addition, we also identified the candidate genes related to other traits, such as the high-altitude hypoxic adaptation

**Table 1** Genetic diversity parameters and the runs of homozygosity (ROH) of the six sheep breeds

Breeds	Abbreviation	ROH number MEAN $\pm$ SD	ROH number	ROH number	ROH number	$F_{ROH}$ MEAN $\pm$ SD	ALL SNPs	
			(0–1 MB)	(1–4 MB)	( $\geq 4$ MB)		$H_e$	$H_o$
Hu Sheep	HS	268.46 $\pm$ 54.41	6904	569	14	0.0528 $\pm$ 0.0199	0.3233	0.2282
Liangshan Black Sheep	LB	510.46 $\pm$ 79.46	8842	784	38	0.1049 $\pm$ 0.0393	0.2835	0.2395
Oula Sheep	OL	238.22 $\pm$ 52.38	2066	69	0	0.0389 $\pm$ 0.0100	0.2283	0.2205
White Xizang Sheep	WX	190.60 $\pm$ 62.62	1834	63	0	0.0310 $\pm$ 0.0117	0.2175	0.2113
Bamei Mutton Sheep	BM	235.50 $\pm$ 112.25	2152	197	0	0.0458 $\pm$ 0.0249	0.2266	0.1828
Poll Dorset Sheep	PD	202.20 $\pm$ 89.26	1855	163	1	0.0393 $\pm$ 0.0210	0.2248	0.1959



**Fig. 1** Population genetic analysis of 6 breeds. **(A)** Principal component analysis of 86 individuals. **(B)** Population genetic structure of sheep. **(C)** Genome-wide distribution of nucleotide diversity of each breed. The points which were on the outside the of whiskers showed outliers. **(D)** Neighbor-joining (NJ) trees of sheep based on whole-genome SNPs

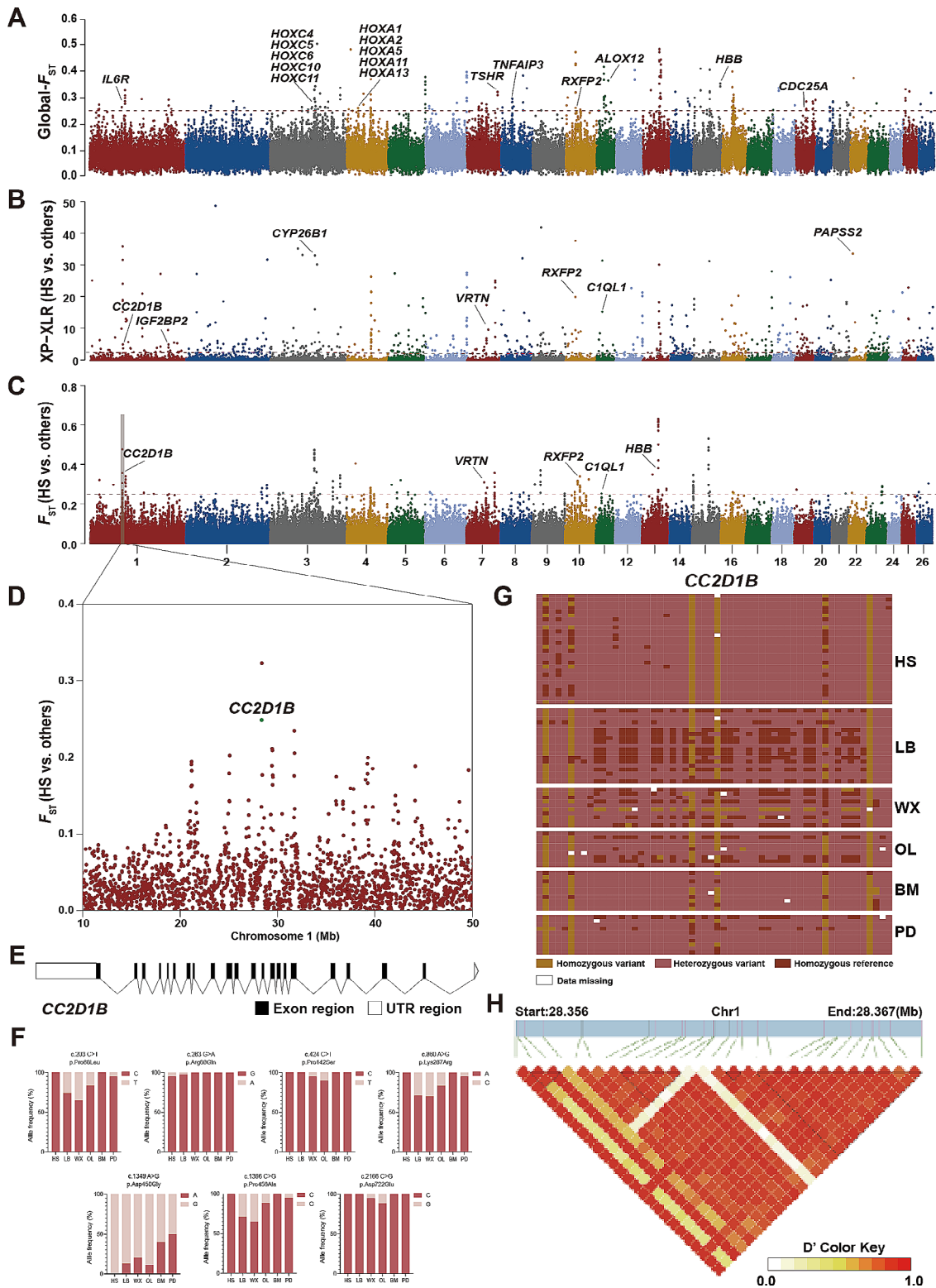
(*HBB*) and regulation of horn growth and development (*RXFP2*).

### Shared selected and candidate genes in Hu sheep

For screening out the associated genetic variation, we estimated the  $F_{ST}$  values and XP-CLR scores to investigate specific genomic regions associated with HS, in comparison to a collective group of five other breeds: LB, WX, OL, PD, and BM. The windows selected as potentially positively selected regions were those simultaneously with both  $F_{ST}$  values > 0.25 and XP-CLR scores > 2.5. A total of 17 regions, encompassing 39 genes, were found to have undergone positive selection in the populations (Fig. 2B, C, Table S4). Genes associated with reproduction (*CCDC103*, *CC2D1B*, *C1QL1*), cell growth (*LIN52*, *CRADD*), growth and development (*VRTN*, *C1QL1*), immune function (*KLRF1*, *MAP3K6*), and other traits were identified. *PAPSS2* was found to be enriched in pathways associated with metabolism, including sulfur metabolism, selenocompound metabolism, and purine

metabolism. What's more, *LIN52* was involved in cellular senescence (Table S5).

*CC2D1B* was observed to have a high  $F_{ST}$  value and XP-CLR scores on chromosome 1 (Fig. 2D), we performed genotype imputation and visualized the data for *CC2D1B*, revealing variations in this gene among different sheep breeds. Meanwhile, the genetic composition of BM and PD was similar, while the genetic composition of HS was distinct from the other five breeds. Additionally, we identified seven missense mutations in *CC2D1B* and found that the allele frequency of each locus differed among breeds (Fig. 2E, F). Among them, the frequency of c.1349 A>G in HS reached 100%, while up to 86.84% and 88.89% in LB and OL, respectively. Visualizing the *CC2D1B* genotype revealed variations among different breeds (Fig. 2G), and haplotype analysis indicated a significant imbalance in linkage within this region (Fig. 2H).



**Fig. 2** Manhattan plot of selective sweeps by  $F_{ST}$  and XP-CLR methods in Hu sheep vs. others. **(A)** Genome-wide distribution of  $F_{ST}$ , which is measured by the average value for each SNP across 5 breeds. **(B)** XP-CLR scores from sheep to two groups: HS, other sheep breeds (LB, WX, OL, BM, PD). **(C)** Fixation index ( $F_{ST}$ ) values from sheep to two groups: HS, other sheep breeds (LB, WX, OL, BM, PD). **(D)** Plot of  $F_{ST}$  windows of chr1 region (10–50 Mb) based on HS and others. **(E)** Structures of *CC2D1B* gene. **(F)** Seven missense mutations of *CC2D1B* gene. **(G)** Haplotype differentiation patterns of *CC2D1B* gene. **(H)** Linkage disequilibrium analysis of SNPs in *CC2D1B* region

### Selection prints in Hu sheep harbored different *FecB* genotypes

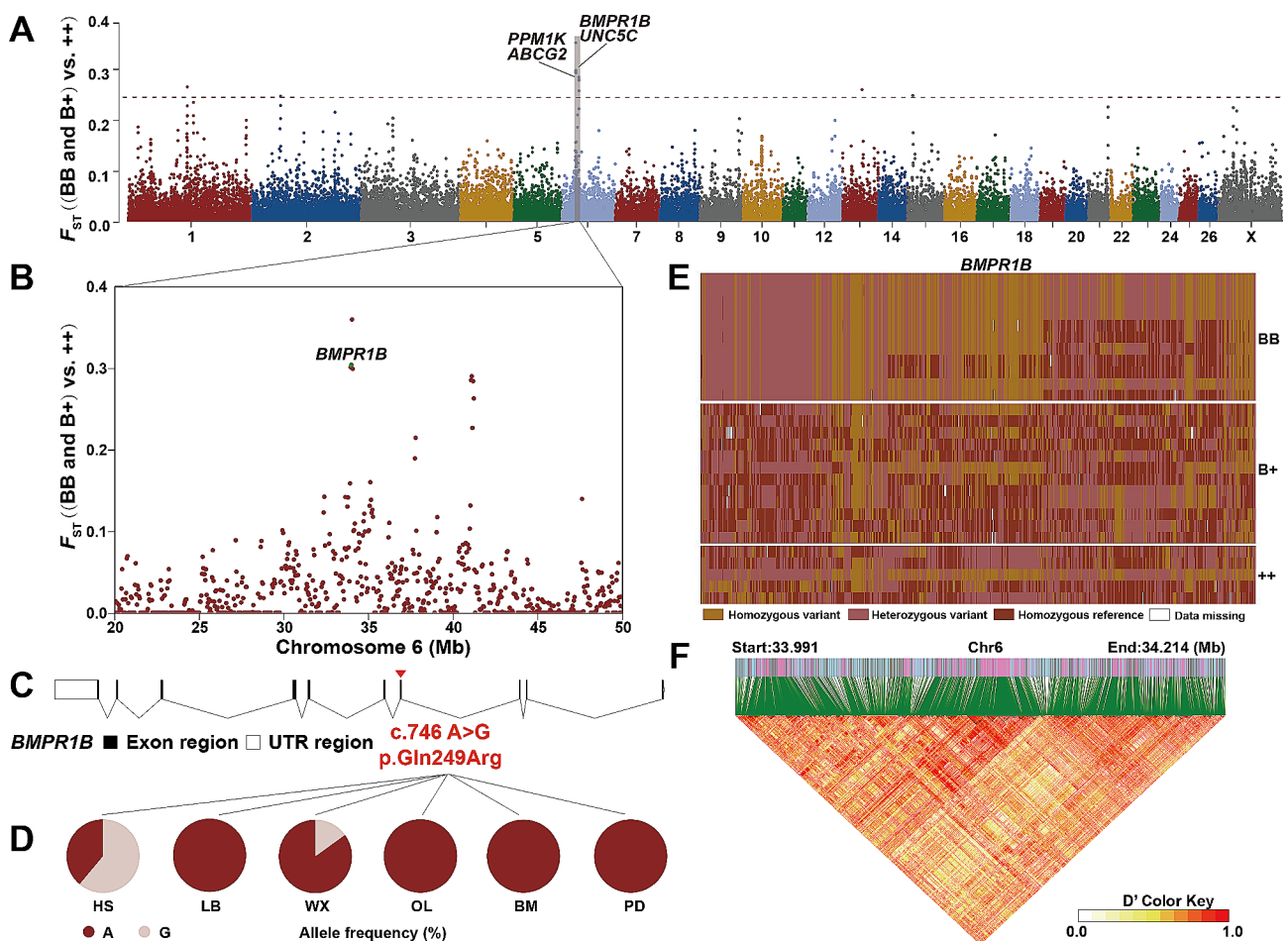
To compare the distribution of *FecB* among different breeds, we examined the *FecB* genotypes of 366 sheep (Table S6). The results showed that there were three genotypes BB (79.9%), B+ (18.6%) and ++ (1.5%) in HS, and only two genotypes B+ (20%) and ++ (80%) were found in WX, and only ++ genotype was detected in LB, OL, BM, and PD.

$F_{ST}$  values were calculated based on different *FecB* genotypes ((BB and B+) vs. ++) to explore critical molecular signals related to the high prolificacy of Hu sheep (Fig. 3A, B). We selected the windows with  $F_{ST} > 0.25$  across the genome and identified three unique autosomal regions with the strongest selective signals containing five candidate genes which included *UNC5C*, *U6*, *ABCG2*, *BMPR1B*, and *PPMIK*. The *BMPR1B* ( $F_{ST} = 0.300978$ ) and *PPMIK* ( $F_{ST} = 0.283851$ ), associated with reproduction traits, were identified on chromosome 6:

33,975,001–34,125,000 bp and 41,100,001–41,250,000 bp, respectively. A significant differentiation was observed in genotype profiles of *BMPR1B* among the three groups with different *FecB* genotypes (Fig. 3E). Many strong linkages were detected in *BMPR1B* (Fig. 3F). Within the *BMPR1B* gene, there was a missense mutation c.746 A>G was identified, which resulted in the substitution of Gln with Arg (Fig. 3C). This specific mutation showed the highest occurrence rate in HS at 61.11%, followed by 15% in WX (Fig. 3D). Furthermore, this mutation was not observed in LB, OL, BM, and PD (Fig. 3D).

### Insight into the novel target genes response to high prolificacy in Hu sheep

Previous studies have confirmed that mutations in the *FecB* gene were completely associated with high reproduction. However, we found some HS ewes carrying the homozygous *FecB* alleles still produced single-offspring. To further determine the significance of genomic



**Fig. 3** Genome-wide annotations during sheep of reproduction. **(A)** Manhattan plot of genome-wide selection signatures associated with *FecB* genotypes in Hu sheep. **(B)** Plot of  $F_{ST}$  windows of chr6 region (20–50 Mb) based on different *BMPR1B* genotypes in Hu sheep. **(C)** Structures of *BMPR1B* gene. **(D)** Allele frequencies of *BMPR1B* gene c.746 A>G. Locations of the SNPs were determined by the generated VCF file. **(E)** Haplotype diversity of a local region of *BMPR1B* (chromosome 6: 33,990,928–34,214,488). **(F)** Linkage disequilibrium analysis of SNPs in *BMPR1B* region

**Table 2** The classification information of different reproduction of Hu sheep

High Reproduction Group (HR)			Low Reproduction Group (LR)		
Individual ID	Litter size	<i>FecB</i> Genotype	Individual ID	Litter size	<i>FecB</i> Genotype
H3	5	BB	H7	1	B+
H6	5	BB	H8	1	B+
H12	4	BB	H9	1	B+
H13	4	BB	H10	1	B+
H15	4	BB	H16	1	BB

divergence, we analyzed and compared the genomes of single-lamb and multi-lamb groups of HS (Table 2), representing high reproduction and low reproduction groups, to identify potential selection imprints. 118 genes were identified as under positive selection with  $F_{ST} > 0.25$  (Fig. 4A). Particularly, *PPPICC*, *CCDC63*, *HSPA2*, and *RHEB* which were associated with reproductive traits, indicating functional importance. In addition, genes related to immune diseases (e.g., *BID*, *SMARCD3*, *ATP6V1E1*, *NFKB2*), energy metabolism (e.g., *HAO1*, *NDUFA7*, *CERS4*, *MTHFD1*), and other factors were also screened (Fig. 4B, Table S7). In KEGG analysis, it was found that *PPPICC* was involved in two reproductive-related biological processes: the oxytocin signaling pathway and oocyte meiosis. *HSPA2* and *RHEB* were identified to participate in the estrogen signaling pathway and thyroid hormone signaling pathway. These pathways played a crucial role in the secretion and regulation of hormones. The KEGG enrichment analysis revealed that candidate genes were significantly enriched in important economic traits, such as reproductive traits, growth and development traits (Fig. 4B, Table S7). Especially, a missense mutation on *CCDC63*, (c.1564 A>G) was identified in all six breeds (Fig. 4D), what's more, haplotype diversity and linkage disequilibrium analysis were performed to detect mutations in gene regions (Fig. 4C, D, E). Simultaneously, allelic frequencies of the c.1564 A>G locus in *CCDC63* were calculated within the 86 individuals of the six breeds (Fig. 4F). The allele frequency in PD was the highest at 45%, while the allele frequency in WX and OL was the lowest at 28.57% (Fig. 4F). KEGG enrichment analysis detected top 30 enriched signaling pathways relevant to candidate genes. Some known pathways related to immunity were found to be significantly enriched (Fig. S3, Table S8).

## Discussion

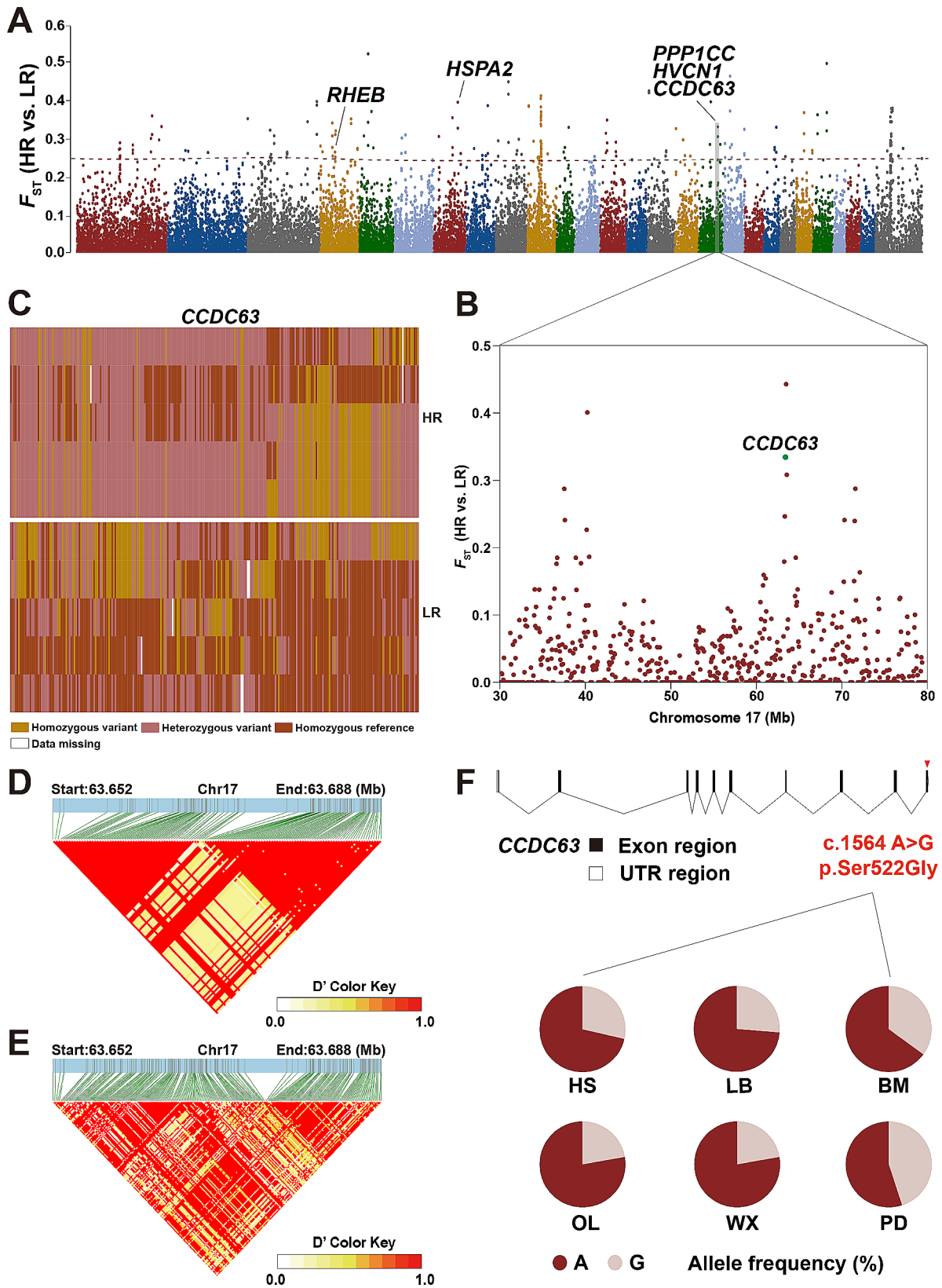
Hu sheep is one of the high reproductive breeds in China, the occurrence of multiple births in Hu sheep is still not clearly understood. Some scholars speculated that it may be influenced by human selection and historical migration patterns [38]. Others suggested that the occurrence

of multiple births in Hu sheep was determined by specific genotypes that can be inherited by subsequent generations [39].

The characteristics of population genetic diversity were essential for assessing the genetic potential of breeds as well as for the utilization and protection of sheep breed resources. The ranking of nucleotide diversity ( $\pi$ ) within different breeds was: HS>LB>OL>WX>PD>BM. One possible explanation for the high nucleotide diversity in HS could be that Hu sheep, with a larger effective population size, originated from the northern Mongolian sheep. Through years of domestication and breeding in a favorable environment, they acquired genetic information from multiple sources [40]. Beyond that, OL and WX exhibited a close genetic relationship, likely due to their proximity geographically and intentional breeding practices [41]. The WX breed was originally bred by Tibetans specifically in the high-altitude environment of the Tibetan Plateau. Known for its superior wool quality, the breed exhibited relatively lower meat productivity. OL, on the other hand, originated from local Tibetan sheep and wild sheep [42, 43]. Genetic diversity, as an essential foundation of biodiversity, was the result of long-term species survival, evolution, and adaptation. This study found that the domestic sheep breed, HS, LB, OL, and WX, exhibited higher genetic diversity, while BM and PD with foreign genetic backgrounds, demonstrated lower genetic diversity. These findings were consistent with the previous studies [44]. Additionally, the ROH analysis indicated that HS and LB, which exhibit longer ROH segments, have undergone adaptive evolution in response to specific environmental pressures, leading to the selection and fixation of certain alleles that confer an advantage [45]. In contrast, shorter ROH segments in OL, WX, BM, and PD may be related to their population history and lower levels of inbreeding [46]. A lower  $F_{ROH}$  value suggests a lower degree of inbreeding and potentially a more diverse gene pool [47]. The high  $F_{ROH}$  and low  $He$  in LB may suggest a need for strategies to increase genetic diversity and reduce inbreeding to avoid potential issues such as inbreeding depression [46]. On the other hand, the low  $F_{ROH}$  and high  $He$  in WX may be beneficial for maintaining a stable population with a rich genetic background [45].

The molecular mechanisms underlying the high prolificacy trait in Hu sheep have not been extensively studied, and so far, no other major molecular markers have been identified except for *FecB*. Genome selection signals have been used to identify genomic regions and genes associated with the reproductive traits of Hu sheep. In our study, a set of genes has been identified through global  $F_{ST}$  analysis in six breeds (Fig. 2A, Table S2). As previously reported, one of the functions of *TSHR* was to regulate the development of the thyroid gland and





**Fig. 4** Selection imprints of different litter sizes. **(A)** Identification of candidate genes related to litter size of Hu sheep. **(B)** Plot of  $F_{ST}$  windows of chr17 region (30–80 Mb) based on HR vs. LR in Hu sheep. **(C)** Haplotype diversity of a local region of *CCDC63* (chromosome 17: 63,652,194–63,687,626) in HR and LR. **(D)** Linkage disequilibrium analysis of SNPs in *CCDC63* gene region of HR group. **(E)** Linkage disequilibrium analysis of SNPs in *CCDC63* gene region of LR group. **(F)** Structures of *CCDC63* gene and allele frequencies of *CCDC63* gene

the secretion of thyroid hormone, which in turn plays an important role in the seasonal reproduction of mammals [48, 49]. *TSHR* could catalyze cAMP synthesis, as a secondary messenger, cAMP was a major contributor to the process of releasing follicle-stimulating hormone and luteinizing hormone. It was noteworthy that *RARA* participated in signaling pathways associated with reproduction, indicating its potential importance in regulating hormones. *RARA* was further predicted to be a transcription factor specifically involved in hair follicle morphogenesis at different stages [50]. We found a region with the highest  $F_{ST}$  value (0.282032) on chromosome 8. *TNFAIP3*, the gene encoding A20, has been documented to be an anti-inflammatory and immune factor [51] and expressed increased abnormally in many types of tumor cells and tissues, which was closely associated with the progression, therapy and prognosis of cancer [52, 53]. Zammit [54] explored the impact of the *Tnfaip3* I325N variant on limiting fecundity by inducing hormonal imbalance, underscoring the role of the anti-inflammatory enzyme *TNFAIP3* in affecting fecundity. Naturally, further studies are needed to delve deeper into the specific mechanisms by which these genes influence sheep reproduction and the implications of mutations on fertility.

We also analyzed the potential selection signatures of HS vs. others using  $F_{ST}$  and XP-CLR methods, overlapping regions that  $F_{ST} > 0.25$  and XP-CLR scores  $> 2.5$  were selected as candidate regions. According to the reference genome annotation information, there are 39 genes located in these selected chromosome regions. We identified candidate genes associated with immune response (*BCAP29* [55], *ATP8B4* [56]), reproductive traits (*CC2D1B*) [57], teat number (*SYNDIGIL*) [58], and horn phenotype (*RXFP2*) [12]. The upregulation of critical genes during the follicular phase reflects the immune system's role in follicular recruitment, while endocrine changes throughout the estrous cycle regulate immune gene expression, affecting follicular atresia and recruitment [59]. The immune system precisely regulates follicle development and maturation, influenced by the ovarian microenvironment [60]. The co-localization of *CC2D1B* and *DPY19L2* suggested that *CC2D1B* plays a crucial role in the reformation of the sperm nuclear envelope [57]. Li found patient P38 with pyriform-headed sperm showed a *CC2D1B* mutation, elucidating the significant role of *CC2D1B* in sperm-head formation [61]. In the present study, *CC2D1B* was located in a significant selective region (Fig. 2B, C), and the genotype pattern differed between the HS and other sheep breeds, suggesting that *CC2D1B* may play an important role in sheep fertility (Fig. 2F). Luongo [62] reported a case where a man exhibited total sperm immotility due to the *CCDC103* p.His154Pro mutation, which confirmed that *CCDC103*

does indeed have an impact on sperm development. The *VRTN* gene variants exhibited a strong correlation with the number of thoracic vertebrae (TVN) in pigs [63]. The loss-of-function mutations of *CC2D1B* contributed to morphological abnormalities of the sperm head [61]. *SYNDIGIL* has been identified as a candidate gene for teat number in pigs [58]. In a study conducted by Lu et al., it was discovered that *CIQL1* played a crucial role in the regulation of follicle failure in the ovaries. This regulation occurred through a multidimensional collaborative mechanism involving intraovarian and endocrine control. These findings demonstrated that the loss of *CIQL1* led to the failure of ovarian follicles, disrupting their proper development and function [64].

A previous study has revealed a close relationship between the *FecB* gene and ovulation number in sheep [65]. Based on different *FecB* genotypes in Hu sheep, the  $F_{ST}$  analysis has identified *BMPRI1B*, *PPMIK*, *UNC5C*, and *ABCG2* genes (Fig. 3A, B). It is widely known, *BMPRI1B* plays a crucial role in the multiple births effect in sheep [66]. Studies using a *PPMIK*-deficient mouse model and downregulated *PPMIK* in human ovarian granulosa cells have shown that *PPMIK* deficiency results in impaired metabolism of branched-chain amino acids, contributing to polycystic ovary syndrome in females [67]. *ABCG2* was found to be associated with milk production in sheep [68].

Using single-lamb HS as the control group, we investigated the positively selected regions associated with multi-lambs, and then discovered genes related to sperm development. Research has shown that the removal of *CCDC63* leads to infertility in male mice, primarily due to the shortening of flagella. Although *CCDC63* did not participate in the formation of the outer dynein arms, it played a crucial role in sperm production [69]. Male mice lacking both PP1 $\gamma$ 1 and PP1 $\gamma$ 2 due to *PPP1CC* knockout were sterile because of impaired sperm morphogenesis. However, fertility and normal sperm function can be restored by introducing transgenic expression of PP1 $\gamma$ 2 alone in the testis of *PPP1CC* mice [70]. *HVCN1* channels were associated with the cryotolerance of mammalian sperm [71] and played a crucial role in modulating sperm motility, kinematics, and facilitating calcium entry into the sperm head in pigs [72].

In conclusion, our study revealed significant genetic differentiations among six distinct sheep breeds at the whole-genome level. Additionally, we identified a set of genes associated with reproductive performance in Hu sheep and visualized the variations in these genes across different sheep breeds. These findings will contribute to the future identification of candidate genes related to reproduction in sheep and identify high-yield gene variants with greater precision which benefits breeding strategies.

## Abbreviations

SNP	Single nucleotide polymorphism
<i>Ho</i>	Observed heterozygosity
<i>He</i>	Expected heterozygosity
$\pi$	Nucleotide diversity
ROH	Runs of homozygosity
LD	Linkage disequilibrium
KEGG	Kyoto encyclopaedia of genes and genomes
$F_{ST}$	Fixation index
XP-CLR	Cross-population composite likelihood ratio test
PCA	Principal component analysis

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10396-x>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8  
Supplementary Material 9  
Supplementary Material 10  
Supplementary Material 11

## Acknowledgements

We thank Tao Lin (Guang'an Feed Industry Management Station) for her support in data analysis.

## Author contributions

TZ designed this study. TZ, DH, and QZ collected samples. TZ, DH, LL, and SZ analyzed the data. HZ, WZ, SY, LW, and LN participated in methodology. TZ and DH wrote the manuscript. All authors have read and approved the final manuscript.

## Funding

This research was supported by the Open Fund of Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province (cndky-2023-02), the Sichuan Natural Science Foundation of China (2023NSFSC1144), and the Key Research and Development of Liangshan (21KJXC0011).

## Data availability

The datasets generated during the present study are available in the NCBI under BioProject accession number PRJNA1058981.

## Declarations

### Ethics approval and consent to participate

All sampling procedures were approved by the Institutional Animal Care and Use Committee at the College of Animal Science and Technology, Sichuan Agricultural University (Chengdu, China, Dky-2021302158). Informed consent was obtained from the owner of the Yuexi Oriental Agricultural Development Co., LTD, before the sampling of the Hu sheep in this study.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 28 December 2023 / Accepted: 8 May 2024

Published online: 15 May 2024

## References

1. Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, et al. Revealing the history of sheep domestication using retrovirus integrations. *Science*. 2009;324(5926):532–6.
2. Tao L, He X, Jiang Y, Liu Y, Ouyang Y, Shen Y, et al. Genome-wide analyses reveal genetic convergence of prolificacy between goats and sheep. *Genes*. 2021;12(4):480.
3. Castilho AC, Nogueira MF, Fontes PK, Machado MF, Satrapa RA, Razza EM, et al. Ovarian superstimulation using FSH combined with equine chorionic gonadotropin (eCG) upregulates mRNA-encoding proteins involved with LH receptor intracellular signaling in granulosa cells from Nelore cows. *Theriogenology*. 2014;82(9):1199–205.
4. Castillo-Gutierrez D, Hernández-Arteaga LES, Flores-Najera MJ, Cuevas-Reyes V, Vázquez-García JM, Loredo-Osti C, et al. Methionine supplementation during pregnancy of goats improves kids' birth weight, body mass index, and postnatal growth pattern. *Biology*. 2022;11(7):1065.
5. Talebi R, Ahmadi A, Afraz F, Sarry J, Woloszyn F, Fabre S. Detection of single nucleotide polymorphisms at major prolificacy genes in the Mehraban sheep and association with litter size. *Annals Anim Sci*. 2018;18(3):685–98.
6. Guan F, Liu SR, Shi GQ, Yang LG. Polymorphism of FecB gene in nine sheep breeds or strains and its effects on litter size, lamb growth and development. *Anim Reprod Sci*. 2007;99(1–2):44–52.
7. Gholizadeh M, Esmaeili-Fard SM. Meta-analysis of genome-wide association studies for litter size in sheep. *Theriogenology*. 2022;180:103–12.
8. Chen S, Guo X, He X, Di R, Zhang X, Zhang J, et al. Insight into pituitary lncRNA and mRNA at two estrous stages in small tail Han Sheep with different FecB genotypes. *Front Endocrinol*. 2021;12:789564.
9. 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(6188):1168–73.
10. Li X, He SG, Li WR, Luo LY, Yan Z, Mo DX, et al. Genomic analyses of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germline innovation. *Genome Res*. 2022;32(9):1669–84.
11. Zhang X, Li W, Liu C, Peng X, Lin J, He S, et al. Alteration of sheep coat color pattern by disruption of ASIP gene via CRISPR Cas9. *Sci Rep*. 2017;7(1):8149.
12. Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, et al. Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *Gigascience*. 2018;7(4):giy019.
13. Lagler DK, Hannemann E, Eck K, Klawatsch J, Seichter D, Russ I, et al. Fine-mapping and identification of candidate causal genes for tail length in the Merinolandschaf breed. *Commun Biology*. 2022;5(1):918.
14. Tao L, Liu YF, Zhang H, Li HZ, Zhao FP, Wang FY, et al. Genome-wide association study and inbreeding depression on body size traits in Qira black sheep (*Ovis aries*). *Anim Genet*. 2021;52(4):560–4.
15. Asadollahpour Nanaei H, Cai Y, Alshawi A, Wen J, Hussain T, Fu WW, et al. Genomic analysis of indigenous goats in Southwest Asia reveals evidence of ancient adaptive introgression related to desert climate. *Zoological Res*. 2023;44(1):20–9.
16. Chantepie L, Bodin L, Sarry J, Woloszyn F, Plisson-Petit F, Ruesche J, et al. Genome-wide identification of a regulatory mutation in BMP15 controlling prolificacy in sheep. *Front Genet*. 2020;11:585.
17. Ahmadi A, Afraz F, Talebi R, Farahavar A, Vahidi SMF. Investigation of GDF9 and BMP15 polymorphisms in mehraban sheep to find the missenses as impact on protein. *Iran J Appl Anim Sci*. 2016;6(4):863–72.
18. Wang H, Wang X, Li T, An X, Yin D, Chen N, et al. Regulation of GDF9 and CDKN1B expression in tibetan sheep testes during different stages of maturity. *Gene Expr Patterns*. 2022;43:119218.
19. Majd SA, Ahmadi A, Talebi R, Koochi PM, Fabre S, Qanbari S. Polymorphism identification in ovine KISS1R / GPR54 gene among pure and crossbreeds of Iranian sheep. *Small Ruminant Res*. 2019;173:23–9.
20. Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol*. 2022;39(2):msab353.
21. Xu YX, Wang B, Jing J, Ma R, Luo Y, Li X, et al. Whole-body adipose tissue multi-omic analyses in sheep reveal molecular mechanisms underlying local adaptation to extreme environments. *Commun Biology*. 2023;6(1):159.

22. Liu Z, Tan X, Wang J, Jin Q, Meng X, Cai Z, et al. Whole genome sequencing of Luxi Black Head sheep for screening selection signatures associated with important traits. *Anim Bioscience*. 2022;35(9):1340–50.
23. Lv X, Chen W, Wang S, Cao X, Yuan Z, Getachew T, et al. Whole-genome resequencing of Dorper and Hu sheep to reveal selection signatures associated with important traits. *Anim Bioscience*. 2023;34(7):3016–26.
24. Zhu M, Yang Y, Yang H, Zhao Z, Zhang H, Blair HT, et al. Whole-genome resequencing of the native sheep provides insights into the microevolution and identifies genes associated with reproduction traits. *BMC Genomics*. 2023;24(1):392.
25. Yao Y, Pan Z, Di R, Liu Q, Hu W, Guo X, et al. Whole genome sequencing reveals the effects of recent artificial selection on litter size of Bamei mutton sheep. *Animals*. 2021;11(1):157.
26. Zhang Y, Xue X, Liu Y, Abied A, Ding Y, Zhao S, et al. Genome-wide comparative analyses reveal selection signatures underlying adaptation and production in Tibetan and Poll Dorset sheep. *Sci Rep*. 2021;11(1):2466.
27. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
29. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
30. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet*. 2008;83(3):359–72.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
32. Alexander JN, Kenneth L. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.
33. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
34. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35(10):1786–8.
35. Wang W, Zhang X, Zhou X, Zhang Y, La Y, Zhang Y, et al. Deep Genome Resequencing Reveals Artificial and Natural Selection for Visual Deterioration, Plateau Adaptability and High Prolificacy in Chinese Domestic Sheep. *Front Genet*. 2019;10:300.
36. Hanning W, Liang Z, Yanbing D, Lingbo M, Cheng J, Hui L, et al. Whole-genome resequencing reveals domestication and signatures of selection in Ujimqin, Sunit, and Wu Ranke Mongolian sheep breeds. *Anim Bioscience*. 2022;35(9):1303–13.
37. Dong SS, He WM, Ji JJ, Zhang C, Guo Y, Yang TL. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief Bioinform*. 2021;22(4):bbaa227.
38. Zhao YX, Yang J, Lv FH, Hu XJ, Xie XL, Zhang M, et al. Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in East Asia. *Mol Biol Evol*. 2017;34(9):2380–95.
39. Chong Y, Liu G, Jiang X. Effect of BMPRII gene on litter size of sheep in China: a meta-analysis. *Anim Reprod Sci*. 2019;210:106175.
40. Lv FH, Peng WF, Yang J, Zhao YX. Mitogenomic meta-analysis identifies two phases of migration in the history of Eastern Eurasian sheep. *Mol Biol Evol*. 2015;32(10):2515–33.
41. Wei C, Wang H, Liu G, Zhao F, Kijas JW, Ma Y, et al. Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Sci Rep*. 2016;6:26770.
42. Tian D, Han B, Li X, Liu D, Zhou B, Zhao C, et al. Genetic diversity and selection of Tibetan sheep breeds revealed by whole-genome resequencing. *Anim Bioscience*. 2023;36(7):991–1002.
43. Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, et al. The genome landscape of Tibetan sheep reveals adaptive introgression from argali and the history of early human settlements on the Qinghai-Tibetan plateau. *Mol Biol Evol*. 2019;36(2):283–303.
44. Xia Q, Wang X, Pan Z, Zhang R, Wei C, Chu M, et al. Genetic diversity and phylogenetic relationship of nine sheep populations based on microsatellite markers. *Archives Anim Breed*. 2021;64(1):7–16.
45. Talebi R, Szmatała T, Mészáros G, Qanbari S. Runs of homozygosity in modern chicken revealed by sequence data. *G3 Genes Genomes Genet*. 2020;10(12):4615–23.
46. Gorsen W, Meyermans R, Janssens S, Buys N. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genet Selection Evol*. 2021;53(1):2.
47. Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nat Commun*. 2021;12(1):2972.
48. Chen J, Bi H, Mats EP, Daiki XS, Angela PFP, Mo C, et al. Functional differences between TSHR alleles associate with variation in spawning season in Atlantic herring. *Commun Biology*. 2021;4(1):795–795.
49. Li M, Zhou Q, Pan Y, Lan X, Zhang Q, Pan C, et al. Screen of small fragment mutations within the sheep thyroid stimulating hormone receptor gene associated with litter size. *Animal Biotechnol*. 2023;34(3):658–63.
50. Zhao B, Luo H, He J, Huang X, Chen S, Fu X, et al. Comprehensive transcriptome and methylome analysis delineates the biological basis of hair follicle development and wool-related traits in Merino sheep. *BMC Biol*. 2021;19(1):197.
51. Priem D, van Loo G, Bertrand MJM. A20 and cell death-driven inflammation. *Trends Immunol*. 2020;41(5):421–35.
52. Gao M, Li X, Yang M, Feng W, Lin Y, He T. TNFAIP3 mediates FGFR1 activation-induced breast cancer angiogenesis by promoting VEGFA expression and secretion. *Clin Transl Oncol*. 2022;24(12):2453–65.
53. Li RS, Zhang Y, Wu B, He M, Shan Z, et al. Clinical and basic evaluation of the effects of upregulated TNFAIP3 expression on colorectal cancer. *Dis Markers*. 2022;2022:1263530.
54. Zammit NW, McDowell J, Warren J, Muskovic W, Gamble J, Shi YC, et al. TNFAIP3 reduction-of-function drives female infertility and CNS inflammation. *Front Immunol*. 2022;13:811525.
55. Seonggyun H, Emily D, T ME, Andrew S, Elliott F, Danli C, et al. Whole-genome sequencing analysis of suicide deaths integrating brain-regulatory eQTLs data to identify risk loci and genes. *Mol Psychiatry*. 2023;28(9):3909–19.
56. Elena L, Lara B, Shervin BPA, P A, Norberto SC. Analysis of ATP8B4 F436L missense variant in a large systemic sclerosis cohort. *Arthritis Rheumatol*. 2017;69(6):1337–8.
57. Ventimiglia LN, Cuesta-Geijo MA, Martinelli N, Caballe A, Macheboeuf P, Miguez N, et al. CC2D1B coordinates ESCRT-III activity during the mitotic reformation of the nuclear envelope. *Dev Cell*. 2018;47(5):547–e563546.
58. Verardo LL, Silva FF, Lopes MS, Madsen O, Bastiaansen JW, Knol EF, et al. Revealing new candidate genes for reproductive traits in pigs: combining bayesian GWAS and functional pathways. *Genet Selection Evol*. 2016;48:9.
59. Talebi R, Ahmadi A, Afraz F. Analysis of protein-protein interaction network based on transcriptome profiling of ovine granulosa cells identifies candidate genes in cyclic recruitment of ovarian follicles. *J Anim Sci Technol*. 2018;60:11.
60. Talebi R, Ahmadi A, Afraz F, Sarry J, Plisson-Petit F, Genêt C, et al. Transcriptome analysis of ovine granulosa cells reveals differences between small antral follicles collected during the follicular and luteal phases. *Theriogenology*. 2018;108:103–17.
61. Li Y, Wang Y, Wen Y, Zhang T, Wang X, Jiang C, et al. Whole-exome sequencing of a cohort of infertile men reveals novel causative genes in teratozoospermia that are chiefly related to sperm head defects. *Hum Reprod*. 2021;37(1):152–77.
62. Luongo FP, Luddi A, Ponchia R, Ferrante R, Di Rado S, Paccagnini E, et al. Case report: the CCDC103 variant causes ultrastructural sperm axonemal defects and total sperm immotility in a professional athlete without primary ciliary dyskinesia. *Front Genet*. 2023;14:1062326.
63. Duan Y, Zhang H, Zhang Z, Gao J, Yang J, Wu Z, et al. VRTN is required for the development of thoracic vertebrae in mammals. *Int J Biol Sci*. 2018;14(6):667–81.
64. Lu X, Ding F, Chen Y, Ke S, Yuan S, Qiu H, et al. Deficiency of C1QL1 reduced murine ovarian follicle reserve through intraovarian and endocrine control. *Endocrinology*. 2022;163(6):bqac048.
65. Davis GH, Balakrishnan L, Ross IK, Wilson T, Galloway SM, Lumsden BM, et al. Investigation of the Booroola (FecB) and inverdale (FecX(l)) mutations in 21 prolific breeds and strains of sheep sampled in 13 countries. *Anim Reprod Sci*. 2006;92(1–2):87–96.
66. Akhatayeva Z, Bi Y, He Y, Khan R, Li J, Li H, et al. Survey of the relationship between polymorphisms within the BMPRII B gene and sheep reproductive traits. *Animal Biotechnol*. 2023;34(3):718–27.
67. Mu L, Ye Z, Hu J, Zhang Y, Chen K, Sun H, et al. PPM1K-regulated impaired catabolism of branched-chain amino acids orchestrates polycystic ovary syndrome. *EBioMedicine*. 2023;89:104492.

68. Signe H, Burren H, Ammann A, Drögemüller P, Flury C. Runs of homozygosity and signatures of selection: a comparison among eight local Swiss sheep breeds. *Anim Genet.* 2019;50(5):512–25.
69. Young SA, Miyata H, Satouh Y, Kato H, Nozawa K, Isotani A, et al. CRISPR/Cas9-mediated rapid generation of multiple mouse lines identified CCDC63 as essential for spermiogenesis. *Int J Biol Sci.* 2015;16(10):24732–50.
70. Dudiki T, Joudeh N, Sinha N, Goswami S, Eisa A, Kline D, et al. The protein phosphatase isoform PP1γ1 substitutes for PP1γ2 to support spermatogenesis but not normal sperm function and fertility†. *Biol Reprod.* 2019;100(3):721–36.
71. Tyler D, Susannah V. Development to blastocyst is impaired when intracytoplasmic sperm injection is performed with abnormal sperm from infertile mice harboring a mutation in the protein phosphatase 1cγ gene. *Biol Reprod.* 2003;68(4):1470–6.
72. Yeste M, Llavenera M, Mateo-Otero Y, Catalán J, Bonet S, Pinart E. HVCN1 channels are relevant for the maintenance of sperm motility during in vitro capacitation of pig spermatozoa. *Int J Biol Sci.* 2020;21(9):3255.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.