Genome-wide association studies of body size traits in Tibetan sheep

Dehui Liu[1,2,4†], Xue Li[1†], Lei Wang[3], Quanbang Pei[3], Jincai Zhao[3], De Sun[6], Qianben Ren[3], Dehong Tian[1], Buying Han[1,2], Hanjing Jiang[5], Wenkui Zhang[3], Song Wang[1,2], Fei Tian[1], Sijia Liu[1] and Kai Zhao[1*]

Abstract

Background Elucidating the genetic variation underlying phenotypic diversity will facilitate improving production performance in livestock species. The Tibetan sheep breed in China holds significant historical importance, serving as a fundamental pillar of Qinghai's animal husbandry sector. The Plateau-type Tibetan sheep, comprising 90% of the province's population, are characterized by their tall stature and serve as the primary breed among Tibetan sheep. In contrast, Zhashijia sheep exhibit larger size and superior meat quality. These two species provide an excellent model for elucidating the genetic basis of body size variation. Therefore, this study aims to conduct a comprehensive genome-wide association study on these two Tibetan sheep breeds to identify single nucleotide polymorphism loci and regulatory genes that influence body size traits in Tibetan sheep.

Result In this study, the phenotypic traits of body weight, body length, body height, chest circumference, chest depth, chest width, waist angle width, and pipe circumference were evaluated in two Tibetan sheep breeds: Plateau-type sheep and Zhashijia Tibetan sheep. Whole genome sequencing generated 48,215,130 high-quality SNPs for genome-wide association study. Four methods were applied and identified 623 SNPs significantly associated with body size traits. The significantly associated single nucleotide polymorphisms identified in this study are located near or within 111 candidate genes. These genes exhibit enrichment in the cAMP and Rap1 signaling pathways, significantly affecting animal growth, and body size. Specifically, the following genes were associated: *ASAP1*, *CDK6*, *FRYL*, *NAV2*, *PTPRM*, *GPC6*, *PTPRG*, *KANK1*, *NTRK2* and *ADCY8*.

Conclusion By genome-wide association study, we identified 16 SNPs and 10 candidate genes associated with body size traits in Tibetan sheep, which hold potential for application in genomic selection breeding programs in sheep. Identifying these candidate genes will establish a solid foundation for applying molecular marker-assisted selection in sheep breeding and improve our understanding of body size control in farmed animals.

Keywords Tibetan sheep, Re-sequencing, GWAS, Body size traits, SNP

[†]Dehui Liu and Xue Li contributed equally to this work.

*Correspondence:
Kai Zhao
zhaokai@nwipb.cas.cn
[1]Qinghai Provincial Key Laboratory of Animal Ecological Genomics, Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, No. 23 Xinning Road, Xining, Qinghai 810001, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Qinghai Sheep Breeding and Promotion Service Center, Gangcha, Qinghai 812300, China
[4]Branch of Animal Husbandry and Veterinary of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161005, China
[5]Qinghai Livestock and Poultry Genetic Resources Protection and Utilization Center, Xining 810000, China
[6]Animal Husbandry and Veterinary Station of Huzhu County of Qinghai Province, Huzhu, Qinghai 810500, China

Liu *et al. BMC Genomics*   (2024) 25:739

Page 2 of 16

Background

Sheep are essential farm animals, with a wide distribution ranging from the Tibetan Plateau to the eastern regions of China. Based on their geographic distribution and genetic relationships, Chinese sheep can be classified into three primary lineages: Tibetan, Mongolian, and Kazakh [1]. Tibetan sheep (*Ovis aries*) exhibit robust adaptability to the alpine pastoral environment and harsh feeding and management conditions and maintain stable genetic properties [2]. As a crucial livestock species in pastoral regions, Tibetan sheep provide essential resources such as meat, milk, fur, and fuel for production and livelihoods, but also is intertwined with social and cultural aspects [3–6]. Due to the absence of systemic selection and breeding, Tibetan sheep showed evident limitations on production traits, such as a low reproductive rate, slow growth, and poor meat production. Zhashijia sheep are an ecological type of Tibetan sheep in Qinghai, which was formed through continuous crossbreeding between wild disc sheep and local sheep and years of natural selection and natural conditions. The Zhashijia family sheep is considered a plateau-type Tibetan sheep. Still, the two breeds differ significantly in body size, and Zhashijia is higher than the plateau type. Both of these two species are Tibetan sheep resources that have been developed through long-term breeding and natural confinement in the pastoral areas of Qinghai Province. However, they exhibit distinct differences in phenotypic characteristics and production traits. Therefore, elucidating the genetic basis underlying these phenotypic variations and identifying novel genes associated with essential characteristics will contribute to the enrichment of genetic resources for superior germplasm development and enhancement of growth performance in sheep.

Body size traits are widely recognized as crucial growth and health indicators, which are significantly influenced by feeding, management, and environmental factors [7, 8]. The noteworthy point is that body size exhibits moderate to high heritability [9–11]. Mature body size has been extensively investigated in humans, cattle, and other domestic animals. However, genomic bases underlying body size variation in sheep remain elusive [9, 12]. Since 1996, Risch et al. [13] recognized the limitations of conventional research methods, prompting the emergence of genome-wide association study as a novel approach. With advancements in high-density SNP microarrays, statistical methodologies, and complementary technologies, GWAS has been widely adopted in sheep genetic breeding programs [14–17]. Zhang [18] employed the SNP50K chip for genotyping 329 sheep from three breeds and identified *PFKFB4*, *TRHED*, *MEF2B*, *PLA2G6*, *RARB*, *OXSM*, *LRRC2*, *ADK*, and *SHISA9* candidate genes were significantly associated with pre-and post-weaning weight gain, weight at six months of age, chest circumference, and calf circumference. Subsequently, among these candidate genes, Zhang [19] further validated the *TRHED* and *MEF2B* genes [18] in Uzhumqin sheep. The study found a significant correlation between these two genes and weight and chest circumference at four months. Ghasemi [20] identified three significant SNPs on chromosome 1 associated with birth weight in 132 Lori-Bakhtiari sheep. These SNPs were predominantly located within the Member RAS Oncogene Family (*RAB6B*) and GRB10 interacting GYF protein 2 (*GIGYF2*) genes, suggesting their potential as candidate genes influencing sheep birth weight. In Baluchi sheep, GWAS revealed that five SNPs were significantly associated with the body weight of an eight-month-old lamb, and one SNP was located within the 50 Kb region near the *MTPN* gene that plays a crucial role in muscle growth and development [21]. Li [22] conducted GWAS on Tibetan sheep and 328 hybrid offspring, identifying ten candidate genes significantly associated with body size traits. These studies have yielded valuable genetic markers that can be utilized in sheep breeding.

Advancements in sequencing technology have propelled the advancement of genomics, with the sequencing of the sheep genome unveiling pivotal functional genes linked to traits such as meat quality, body size, and growth. Furthermore, it is feasible to pinpoint loci responsible for phenotypic differences caused by variations, which play a pivotal role in guiding the analysis of animal genetic mechanisms.

This study aimed to conduct a genome-wide association study (GWAS) using 11× whole genome resequencing data and body size traits of 193 sheep from plateau-type Tibetan and Zhashijia breeds. The objective was to identify candidate genes and SNP loci that influence body size traits in Tibetan sheep and to discover potential genetic markers for genomic selection in sheep breeding programs.

Results

Descriptive statistics

We conducted descriptive statistics and analyzed the phenotypic data of 193 Tibetan sheep. The results presented in Table 1 and Supplementary Table S1 include the mean, maximum, minimum, standard error, and coefficient of variation for body weight (BW: 40.4 kg), body length (BL: 66.28 cm), body height (BH: 69.66 cm), chest circumference (CC: 91.52 cm), chest depth (CD: 32.01 cm), chest width (CW: 21.52 cm), waist angle width (WW: 13.84 cm) and pipe circumference (PC: 7.38 cm). The coefficient of variation ranged from 6.86 to 21.73%, with the lowest and highest values observed for traits BH and BW, respectively. Finally, we performed correlation analyses for the eight phenotypic traits and found that only PC and WW were not significantly correlated, while

Table 1 Descriptive statistics for the body size traits evaluated

Trait	Mean	Max	Min	SE	CV(%)
BW (kg)	40.04	77.1	21.9	0.63	21.73
BH (cm)	69.66	84	57	0.34	6.86
BL (cm)	66.28	80	41	0.39	8.12
CC (cm)	91.52	107.5	69.5	0.53	8.1
CW (cm)	21.52	30	17	0.15	9.67
CD (cm)	32.01	41	24	0.21	9.06
WW (cm)	13.84	16	10	0.08	7.86
PC (cm)	7.38	9.2	6.2	0.04	7.05

Note BW=body weight; BH=body height; BL=body length; CC=chest circumference; CW=chest width; CD=chest depth; WW=waist angle width; PC=pipe circumference

all other phenotypic characteristics were significantly correlated (Supplementary Table S2).

Whole genome sequencing data

A total of 193 Tibetan sheep were subjected to whole genome sequencing, generating 5.49 Tb of raw reads with an average of 29.15 Gb per sample. The sequencing quality was high, with an average Q20 score of 97.15% and an average Q30 score of 91.15%. The GC content distribution in the 193 samples ranged from 41.65 to 45.78%, indicating no bias in library construction and sequencing (Supplementary Table S3). Compared to the reference genome, the average coverage depth ranged from 193 samples (Supplementary Table S4). Following SNP calling, we identified 48,755,632 SNP sites and classified them into 12 distinct mutation types. Among these mutations, C>T and G>A were the most prevalent mutations, occupying 19.5% and 19.6%, respectively, and A>T, C>G, G>C, and T>A were the most minor frequent mutations, occupying only 3.7% (Supplementary Fig. 1 and Table S5). Annotation results showed that 52.75% and 34.29% of SNPs were in intergenic and intron regions, and only 0.41% represented missense mutations (Supplementary Table S6).

Population genetic diversity analysis

Based on SNPs, a phylogenetic tree was inferred using the Maximum Likelihood (ML) method and revealed a paraphyletic relationship between the PT and ZS populations (Supplementary Fig. 2). In line with the phylogenetic tree, STRUCTURE analysis indicated that when K equaled 2, PT and ZS were not completely genetically distinguished (Fig. 1b and Supplementary Fig. 3). Component Analysis (PCA) showed that PC1 and PC2 accounted for 1.15% and 0.97% of differences, and PT and ZS were separated into two groups along the PC1 axis (Fig. 1a). The decay of linkage disequilibrium (LD) demonstrated that PT had a higher LD rate than ZS, implying greater genetic diversity in the former and higher selection pressure on the latter population (Fig. 1c). These observations indicated that the artificial selection led to the genetic differentiation

between PT and ZS, suggesting the incorporation of population stratification as a covariate in the following GWAS to minimize confounding effects.

Genome-wide association studies of body size traits

The IBS-based kinship matrix indicated that most individuals are unrelated, and some are genetically related in 193 Tibetan sheep (Supplementary Fig. 4), suggesting the consideration of kinship in GWAS. Additionally, we included sex and age as fixed effects to mitigate their influence on specific phenotypes, considering that these variables generally influence body size values.

Based on the criterion of a $-\log_{10}$ (P-value) more significant than 5, BLINK, GLM, MLM, and SUPER models identified 520, 287, 166, and 398 SNPs significantly correlated with body size traits. The four algorithms existed to detect the same SNP loci, so we integrated and analyzed the number of loci detected by each of the four algorithms. It was determined that the four algorithms concurrently identified a total of 623 SNP loci associated with body size traits, including 98 SNPs for body weight, 35 SNPs for body height, 116 SNPs for body length, 41 SNPs for chest circumference, 87 SNPs for chest depth, 56 SNPs for chest width, and 40 SNPs for pipe circumference as well as an additional set of 150 SNPs related to waist angle width.

After gene annotation, 10 candidate genes associated with body size traits were ultimately identified, among which 16 SNP mutations were detected. Of these, 98 variants exhibited significant correlations with body weight (see Table S7 and Fig. 2 for detailed information). These body weight-correlated SNPs were annotated in 20 genes, and the most statistically significant SNP (Chr9: 23,424,482 G>A mutation) occurred in the intron of Arf-GAP with SH3 domain, ankyrin repeat and PH domain 1 (*ASAP1*) gene.

The BH trait was closely associated with 35 SNPs (Fig. 3). Annotation revealed one SNP-annotated gene linked to the body size traits. This specific SNP (Chr2: 69,002,239 C>G mutation) was in the intron of the KN motif and ankyrin repeat domains 1 (*KANK1*) gene.

The Manhattan plot revealed significant associations between 41 SNPs and the CC trait, annotated in 11 genes (Fig. 5). Notably, the most significant SNP within neuron navigator 2 (*NAV2*) gene was identified on Chr21 as a T>G mutation (position: 22,788,649) occurring in an intron. Additionally, one significant SNP (Chr2: 34,862,294 G>A mutation) was annotated in the intron of neurotrophic receptor tyrosine kinase 2 (*NTRK2*) gene. Furthermore, five SNPs were found within the genomic interval of 39,525,865−39,557,828 on chr19. These five mutations (C>T, T>A, G>T, A>G and A>G) are all located in the intronic region of protein tyrosine phosphatase receptor type G (*PTPRG*) gene.
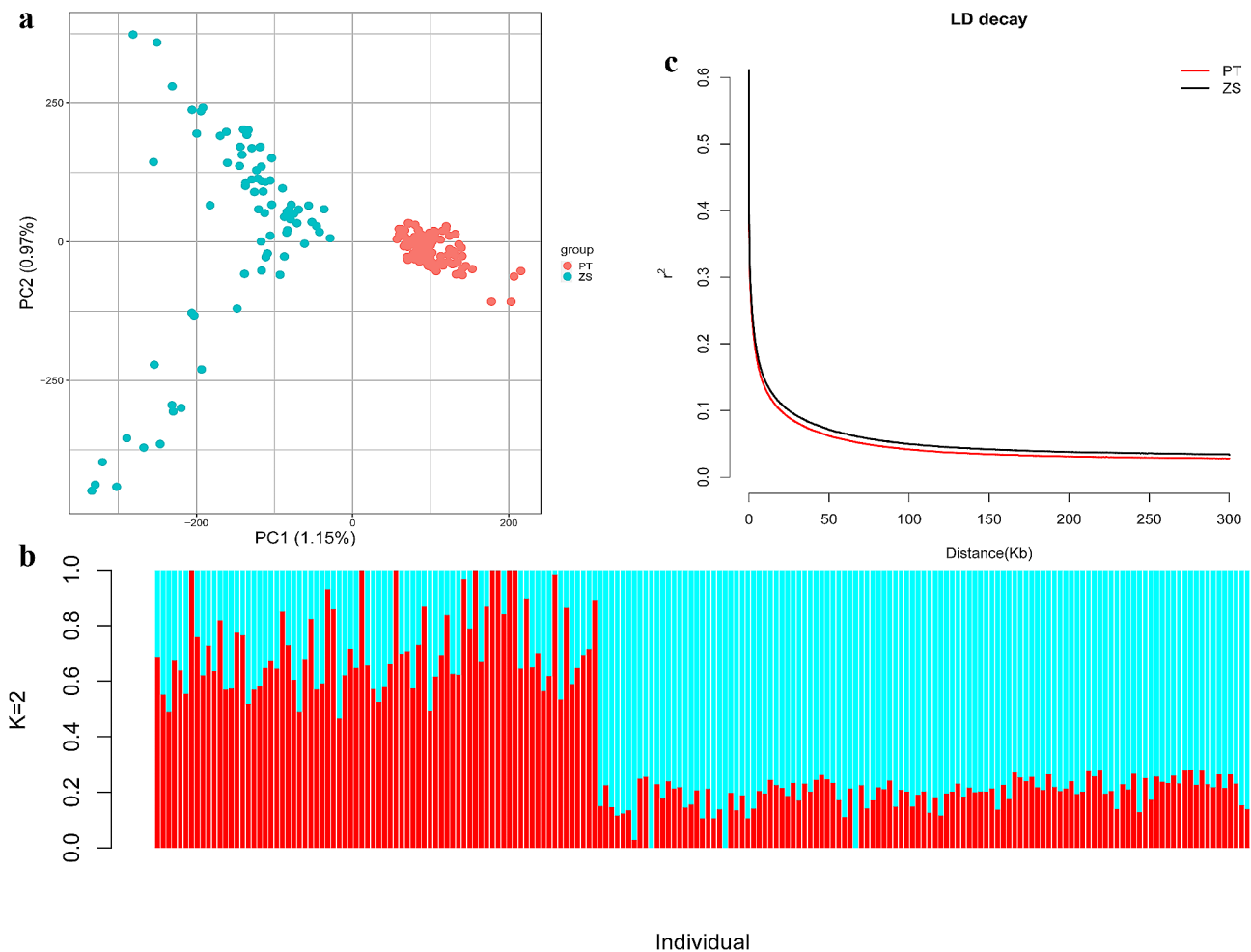
Fig. 1 a: PCA analysis for the two populations of Tibetan sheep. PT indicates, and ZS denotes; **b**: Map of population genetic structure. The red color in the figure represents a subpopulation component, and the blue color represents a subpopulation component. Each vertical bar represents a sample, with 1–78 being ZS and 79–193 being PT. Population genetic structure analysis assumes that several lines originated from K distinct (or widely differing) ancestors and analyses the proportions of each hypothetical ancestral component each line has in its genetic composition. The best results are obtained when K = 2; **c**: The linkage disequilibrium (LD) decay analysis. PT indicates Plateau type Tibetan sheep, and ZS denotes Zhashijia sheep

Statistically, 87 single nucleotide polymorphisms (SNPs) reached genome-wide significance for the CD trait (Fig. 7), and four genes associated with body size traits were identified through gene annotation. Among them, one significant SNP (Chr4: 10,696,280 C > T mutation at the intron) was located within the intron region of cyclin dependent kinase 6 (*CDK6*). Additionally, two significant SNPs (Chr6: 68,113,923 A > G mutation at the intron; Chr9: 23,072,987 T > C mutation at the intron) were annotated as the FRY like transcription coactivator (*FRYL*) and adenylate cyclase 8 (*ADCY8*) genes, respectively. Furthermore, a significant SNP site on chr10 was observed to have a G > A mutation (position: 68,522,087) within its intron region, which was annotated as the glypican 6 (*GPC6*) gene.

According to the Manhattan plot, 150 SNPs were identified as significantly associated with the WW phenotype (Fig. 8). These SNPs were annotated within a set of 21 genes. *CDK6* and protein tyrosine phosphatase receptor type M (*PTPRM*) emerged as potential candidates influencing body size traits among these genes. Notably, a C > T mutation was observed at position 41,029,246 on chromosome 23, representing a significant SNP site within the *PTPRM* gene's intron region.

This study's BL trait was significantly associated with 116 SNPs in 17 genes (Fig. 4). Besides, the CW phenotype demonstrated an association with 56 SNPs that were annotated to 15 genes (Fig. 6). The combined analysis revealed that 40 SNPs, annotated to seven specific genes, were significantly associated with PC traits (Fig. 9). Among these SNP-annotated genes, no genes were found to be associated with body size traits.

The QQ plots revealed a consistent relationship between most loci's actual and theoretical values, indicating the effectiveness of the four methods in hierarchical population control.
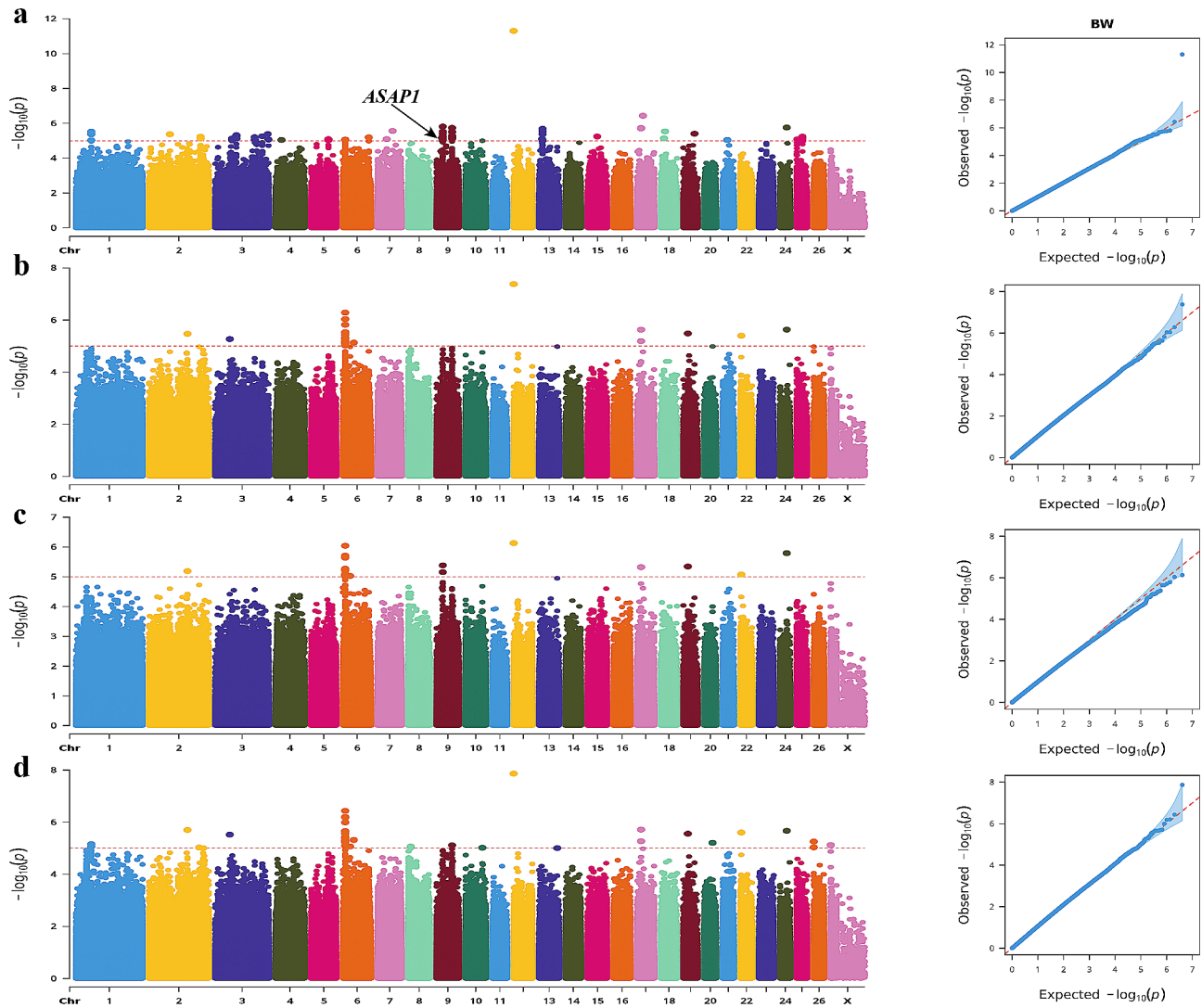
Fig. 2 Manhattan and QQ plots of BW trait. (**a**). BLINK; (**b**) GLM (**c**) MLM (**d**) SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -$\log_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

Enrichment analysis

GO and KEGG enrichment analyses were performed to comprehensively assess the function of candidate genes annotated by SNPs (Supplementary Table S8 and Fig. 10). The eight phenotypes (BW, BH, BL, CC, CD, CW, WW, and PC) were collectively submitted for enrichment analysis. Through GO enrichment analysis, we identified a total of 10 significantly related terms, with genes mainly distributed in the terms "cytoskeleton", "cytoskeletal protein binding", "actin filament-based processes", and "nervous system development". The KEGG enrichment analysis revealed significant gene enrichment in signal transduction and metabolism pathways, including the cAMP and Rap1 signaling pathways, bile secretion, and Insulin secretion. Candidate gene *ASAP1, CDK6, FRYL, NAV2, PTPRM, GPC6, PTPRG, KANK1, NTRK2,* and

ADCY8 were identified to be involved in these function pathways and go terms, suggesting their roles in controlling body size through the regulation of metabolism.

Discussion

Compared to commercial and domestic superior breeds, Tibetan sheep exhibit disadvantages in body size and growth rate. Body size encompasses various intricate traits such as body weight, height, length, chest depth, width, circumference of the chest and pipe, and waist angle width. Enhancing these traits through traditional breeding methods within a limited timeframe proves challenging due to slow progress rates, low selective accuracy for individual characteristics, high input costs, and time-consuming processes. With the rapid advancements in sequencing and statistical methods, genetic
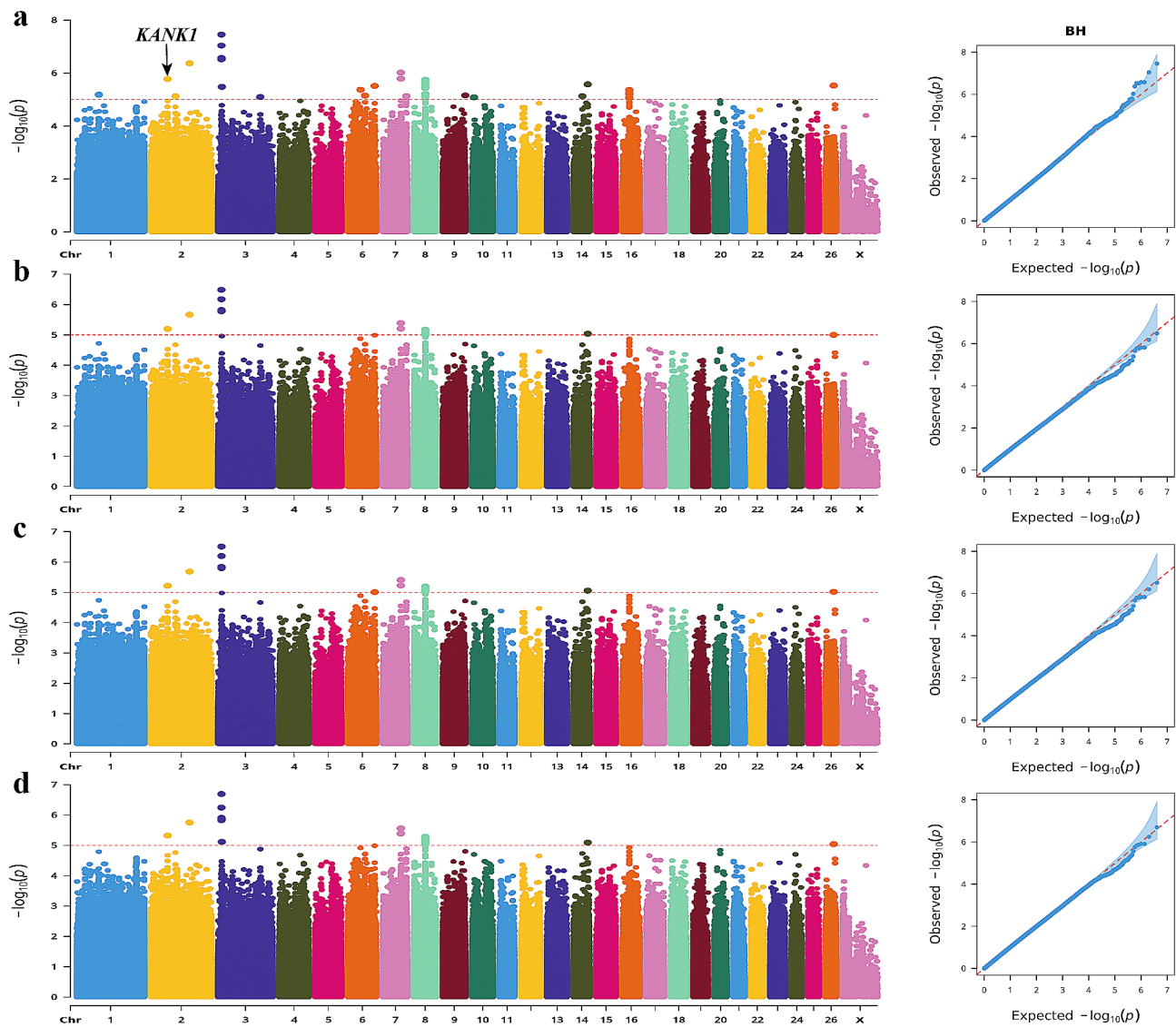
Liu *et al. BMC Genomics*   (2024) 25:739

Page 6 of 16

Fig. 3 Manhattan and QQ plots of BH trait. **a**. BLINK; **b**. GLM **c**. MLM **d**. SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -$\log_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -$\log_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

markers and candidate genes were detected to be associated with body size traits, facilitating the breeding progression. In this study, we conducted genome-wide sequencing and association studies to identify SNPs and candidate genes influencing body size in the Tibetan sheep (PT and ZS breeds), which lays the foundations for the genetic enhancement of growth performance in sheep breeding.

In correlation analyses, group stratification and individual kinship are commonly recognized as the primary factors that affect false-positive associations. To mitigate these influences, the leading analytical tool, GAPIT, with multiple algorithms (GLM, MLM, CMLM, ECMLM, SUPER, FarmCPU, and BLINK), was adopted

for genome-wide association analysis in the current study [23, 24].

Currently, GWAS studies have primarily focused on reproductive and meat quality traits [25, 26], with limited exploration of body size traits in sheep, such as Hulunbeier and Lake sheep. In contrast, Tibetan sheep have garnered more attention for their environmental adaptation capabilities. Therefore, we conducted this study to address this research gap. Zhang et al. discovered a significant association between OAR17_14085599 SNPs and chest circumference in Hulun Buir sheep [27]. Jiang et al. identified five SNPs associated with body height and four SNPs associated with chest circumference in Hu sheep [7]. In our study, we detected 16 noteworthy SNPs at the genomic level: one linked to BW, one related to BH, seven
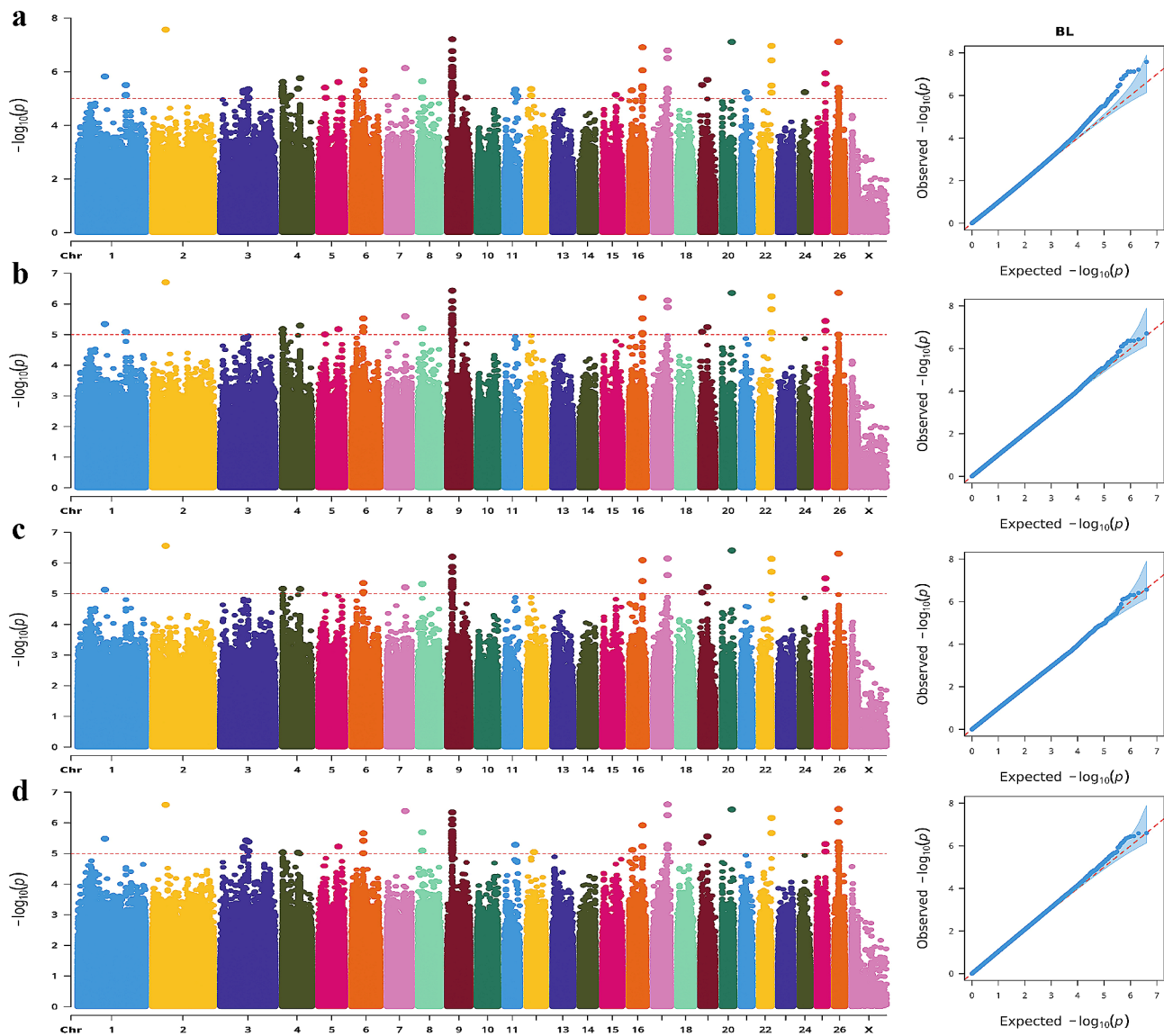
Fig. 4 Manhattan and QQ plots of BL trait.**a**. BLINK; **b**. GLM **c**. MLM **d.** SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log$_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -log$_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

linked to CC, four connected to CD, and three linked to WW. These SNPs are located within intronic regions, which might impact body size traits by regulating the splicing pattern or transcription efficiency. Therefore, we conducted a more comprehensive investigation of these SNP mutation sites and performed genomic annotation on these 16 SNPs. As a result, we identified ten potential candidate functional genes that may exert pivotal roles in the growth, development, and regulation of body size.

This study identified SNPs associated with body weight trait and revealed significant associations between the *ASAP1* gene and growth. Consistent with our result, SNPs in the *ASAP1* gene have been reported to be associated with production and growth traits in cattle [28,

29]. It is also linked to growth, adipocyte differentiation, and the formation of fat accumulation in mice [30]. Furthermore, the *ASAP1* gene has also been identified as a candidate gene for age-specific body weights and plays a crucial role in regulating duck growth rate, growth, and development [31]. After conducting an association analysis of CD trait, we successfully identified the genes *FRYL*, *ADCY8*, and *GPC6*. *FRYL* has been suggested as a potential candidate gene influencing the body weight of broiler chickens, thus indicating its significance in chicken growth [32]. Notably, the *ADCY8* gene not only exhibits a strong association with human obesity but also enhances carcass quality and meat quality in sheep [33, 34]. Carcass weight is a body weight characteristic.
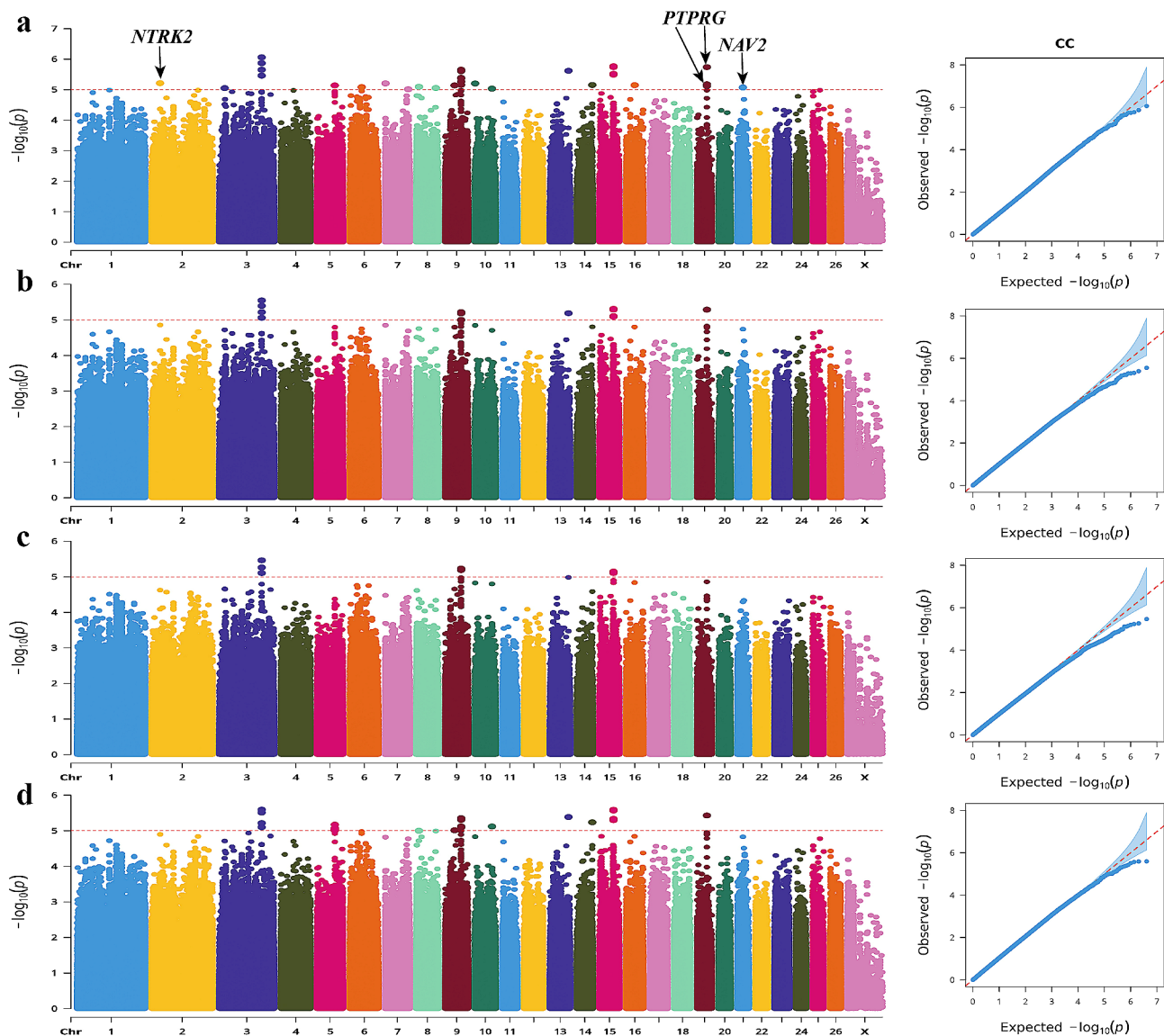
Fig. 5 Manhattan and QQ plots of CC trait. **a**. BLINK; **b**. GLM **c**. MLM **d.** SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log$_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -log$_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

Enrichment analysis revealed that the *ADCY8* gene was significantly enriched in both the cAMP and Rap1 pathways. The cAMP pathway plays a crucial role in regulating adipogenesis, lipid partitioning in white adipose tissue, as well as lipid metabolism in the liver. Additionally, it is involved in central physiological processes such as cell growth, differentiation, gene transcription, protein expression, and metabolism [35–37]. On the other hand, Rap1 signaling has been shown to have a impact on the β-adrenergic signaling pathway and plays an important role in skeletal muscle growth and development, which are essential for animal growth [38, 39]. These findings further support the possibility that the mutation in *ADCY8* gene may influence body size in Tibetan sheep.

The *GPC6* gene has been investigated in mice and pigs, revealing that loss-of-function mutations in this gene lead to reduced stature and shortened limbs in mice [40]. Furthermore, the *GPC6* gene is considered a potential contributor to intramuscular fat deposition in Duroc pigs while exhibiting a significant correlation with the large body size of Sujiang pigs. Additionally, the *GPC6* haplotype plays a role in enhancing the growth performance of Laiwu pigs [41, 42]. The *CDK6* gene was identified to associate with WW and CD traits in Tibetan sheep. A previous study has showed that SNPs in the *CDK6* gene had associations with body length, loin height, and chest circumference in cattle [43]. In addition to domestic animals, this gene has been identified to be associated with
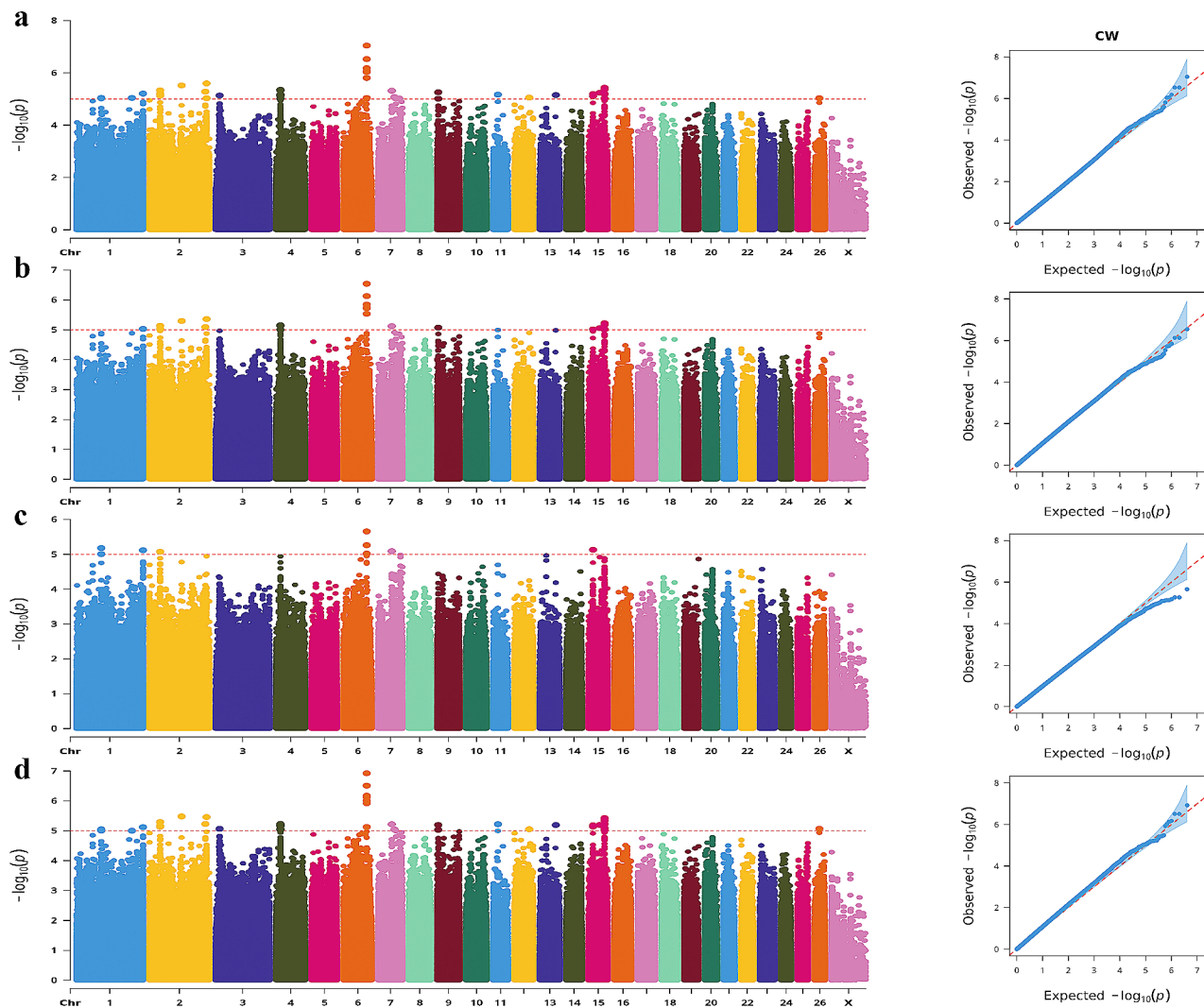
Fig. 6 Manhattan and QQ plots of CW trait. **a**. BLINK; **b.** GLM **c**. MLM **d**. SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log$_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -log$_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

body height and morphological dimensions in humans, suggesting its convergent rwuole in regulating body size [44]. Additionally, the WW trait association analysis revealed the involvement of the *PTPRM* gene, which governs a diverse array of cellular processes encompassing cell growth, differentiation, and signaling cascades. It has been reported that genes involved in cell growth and differentiation experienced selection during domestication and might contribute to an enhanced growth rate in pigs [45]. CC trait association analyses identified the *PTPRG* gene as a potential QTL gene for growth and linear traits [46]. *NAV2* gene exhibited a significant association with mouse height [47]. The genes *GPC6*, *CDK6*, *PTPRM*, *PTPRG*, and *NAV2* have been found to influence body size in various animal species. Therefore, we postulated that these genes might be correlated with phenotypic traits related to body size in Tibetan sheep.

In the current study, we newly identified the significant association between *NTRK2* and *KANK1* genes with CC and BH traits in Tibetan sheep. The *NTRK2* gene regulates human obesity [48, 49]. At the same time, Kank1 has been demonstrated to play a crucial role in the actin cytoskeleton dynamics and is commonly expressed across mammalian tissues. Moreover, actin cytoskeleton dynamics exert significant regulatory control over muscle development, growth, and regeneration by governing fundamental processes such as cell proliferation, differentiation, and morphological changes [50]. Therefore, *NTRK2* and *KANK1* genes may influence body size traits in Tibetan sheep and serve as potential candidates for genetic selection.

Identifying SNPs and candidate genes associated with body size traits will provide valuable resources for enhancing growth performance in Tibetan sheep. In this
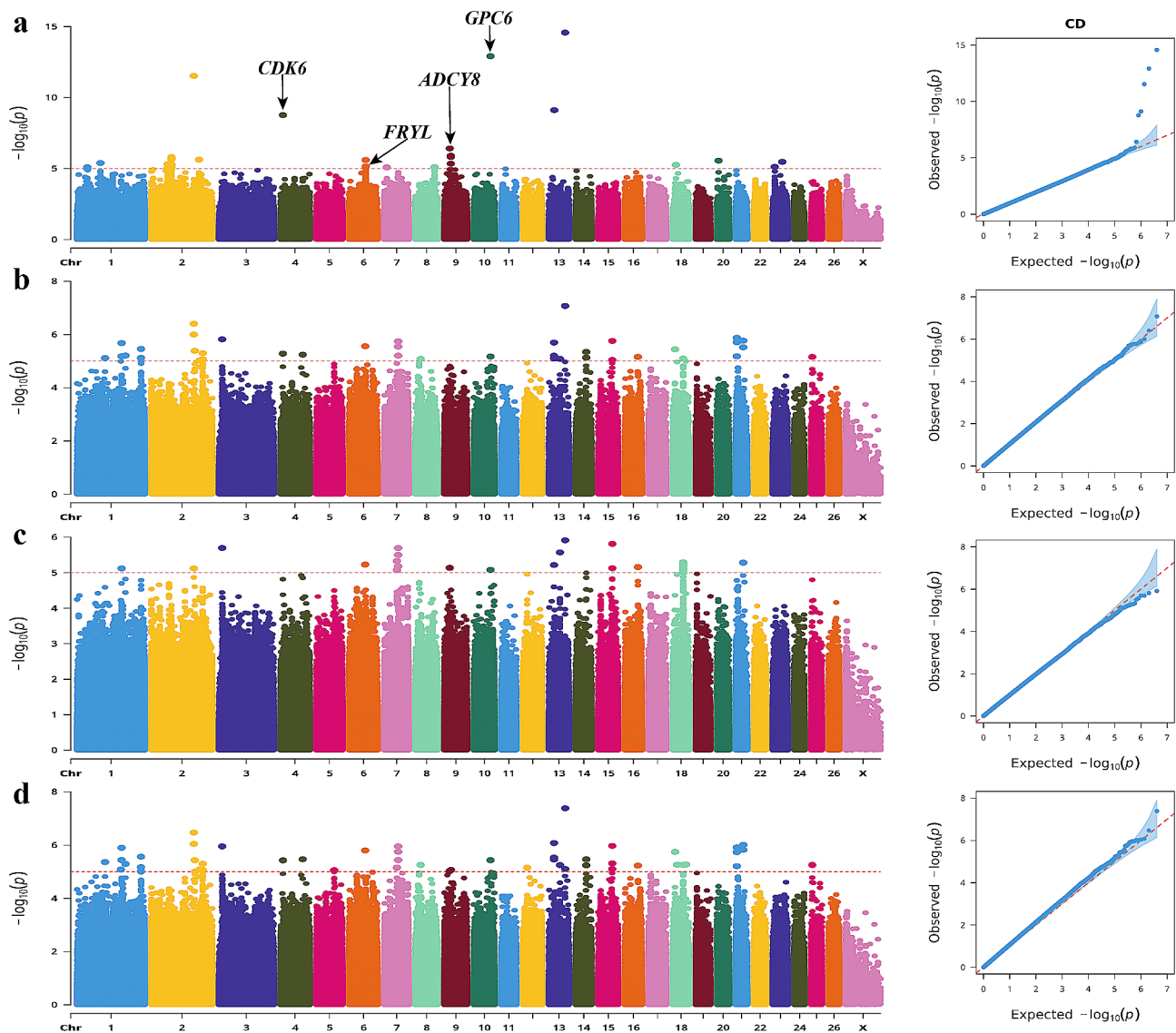
Liu *et al. BMC Genomics*   (2024) 25:739

Page 10 of 16

Fig. 7 Manhattan and QQ plots of CD trait. **a.** BLINK; **b.** GLM **c.** MLM **d.** SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log$_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -log$_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

study, we have identified ten genes (*ASAP1, CDK6, FRYL, NAV2, PTPRM, GPC6, PTPRG, KANK1, NTRK2,* and *ADCY8*) potentially influencing the phenotypic variation in body size of Tibetan sheep. Subsequently, we aim to investigate their transcriptional and protein expression levels to discern any differential regulation and explore the enriched pathways associated with these genes to elucidate the underlying mechanisms governing body size regulation in Tibetan sheep.

Conclusion

Conclusively, we conducted genome-wide sequencing of 193 sheep from two Tibetan sheep breeds (PT and ZS) and performed GWAS using four models to investigate eight phenotypes related to body size. In total, 623 SNPs

were significantly associated with eight body size traits. Based on these SNPs, 111 candidate genes were identified. Enrichment analysis revealed that *ASAP1, CDK6, FRYL, NAV2, PTPRM, GPC6, PTPRG, KANK1, NTRK2,* and *ADCY8* genes were involved in pathways and terms related to growth and development, which might contribute to the regulation of the body size. Considering the findings above collectively, these genes may play crucial roles in determining body size traits in Tibetan sheep. The present study has successfully identified many significant SNPs associated with body size traits in sheep, thereby establishing a robust theoretical foundation for enhancing sheep production performance. Further analyses are required to characterize the functions of these
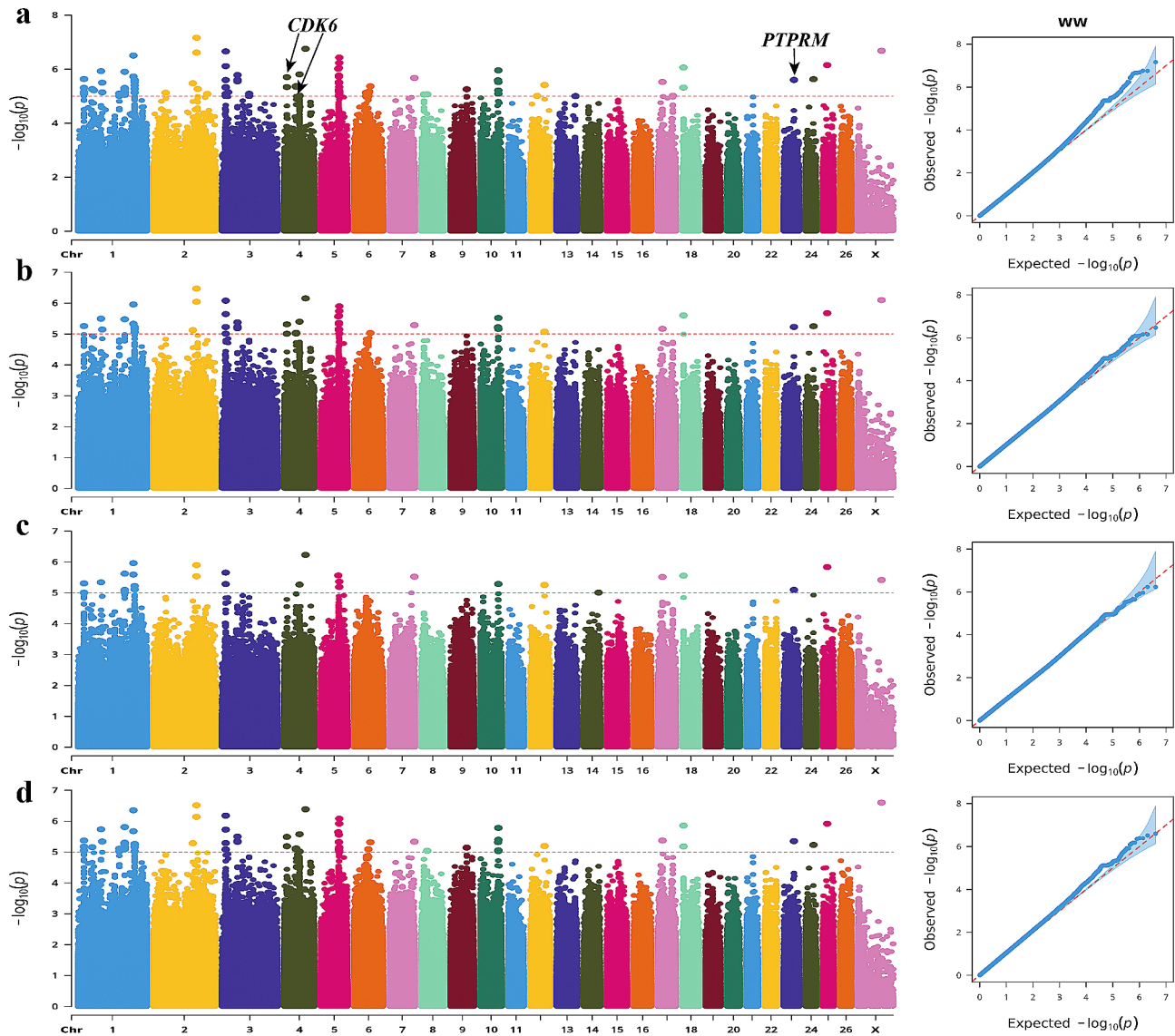
Fig. 8 Manhattan and QQ plots of WW trait. **a**. BLINK; **b**. GLM **c**. MLM **d**. SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -$\log_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -$\log_{10}$ (P-value), and points beyond the threshold range are significantly associated loci.

genes to resolve the genetic mechanisms affecting sheep phenotypes.

Materials and methods

Animal ethics

The animal-related procedures were strictly conducted by the guidelines and regulations established by the Institutional Animal Care and Use Committee of the Northwest Institute of Plateau Biology, Chinese Academy of Sciences (Approval No. NWIPB2023015). The animal owners provided informed consent to include all animals in the study.

Sample collection and phenotypic recording

The ear tip tissues were collected from 193 Tibetan sheep (1 to 3.5 years), comprising 115 plateau-type sheep (56 rams and 59 ewes) and 78 Zhashijia Tibetan sheep (20 rams and 58 ewes). These samples were obtained from the Qinghai Sheep Breeding and Promotion Service Center in Gangcha County and a cooperative in Qumalai County, Qinghai Province, China. The ear tip tissues were preserved in 75% alcohol at -20 °C for subsequent DNA extraction. Eight phenotypes of body size traits were recorded in 193 Tibetan sheep, including body weight, body length, body height, chest circumference, chest depth, chest width, waist angle width, and pipe
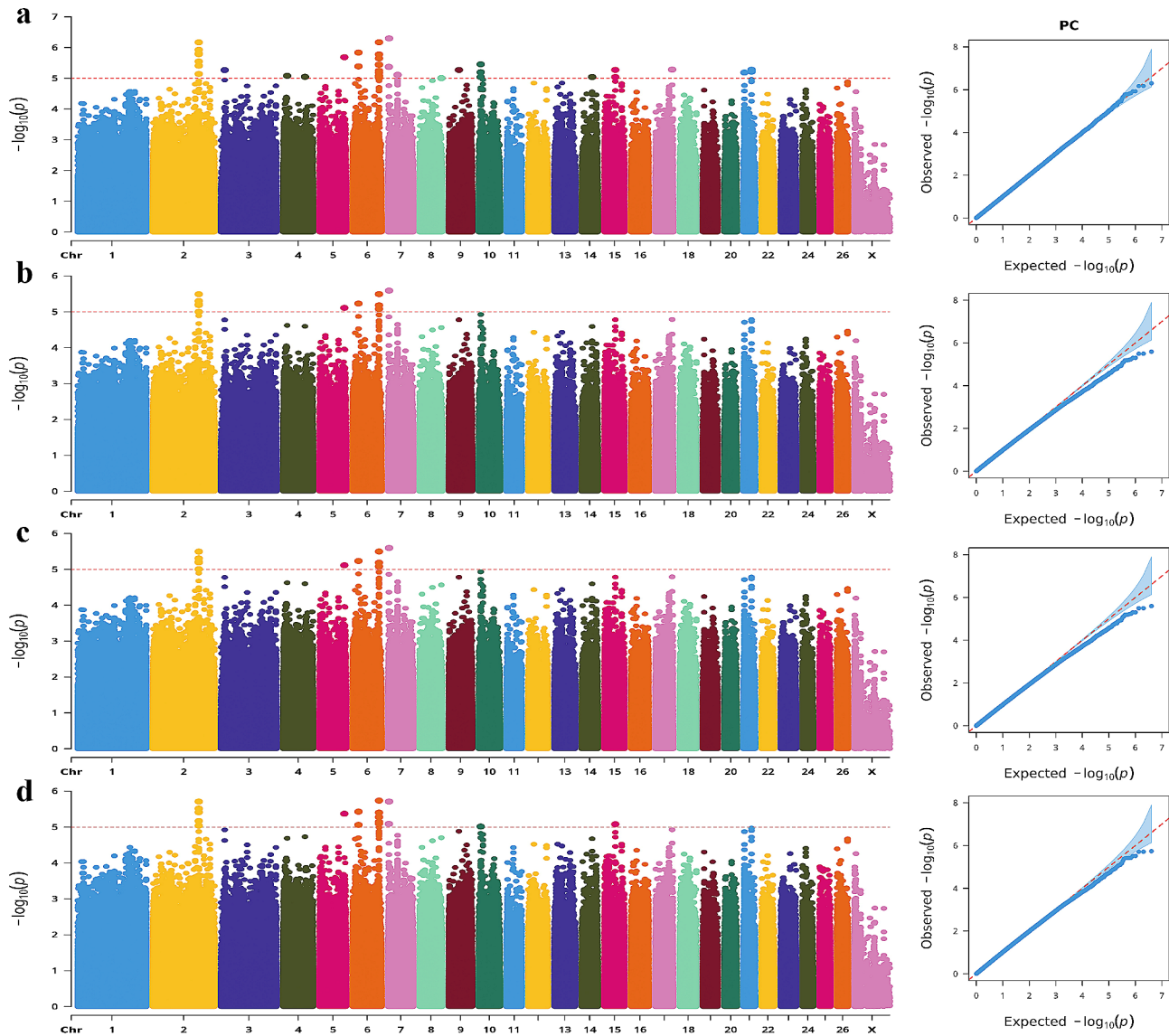
Fig. 9 Manhattan and QQ plots of PC trait. **a**. BLINK; **b**. GLM **c**. MLM **d**. SUPER. The horizontal coordinates are all chromosomes, and the vertical coordinates are loci -log$_{10}$ (P-value); four analysis models were used (**a**, **b**, **c**, and **d**) for GWAS analysis, the red dashed line is the threshold for -log$_{10}$ (P-value), and points beyond the threshold range are significantly associated loci

circumference. We utilized the PASW Statistics 18 software to establish correlations among the eight phenotypic traits.

Whole-genome sequencing

Genomic DNA was extracted and purified from 193 Tibetan sheep samples using a tissue kit (BGI-Shenzhen, China). DNA of 0.5 µg was used for library construction based on the manufacturer's description, with an insert size of approximately 300 bp. After PCR and purification, the sequencing library was sequenced in the DNBSEQ platform (BGI-Shenzhen, China).

We employed the software SOAPnuke (version: 2.2.1) [51] for quality control and filtration of the raw data.

Subsequently, BWA's mem algorithm (version: 0.7.17-r1188) [52] was utilized to align clean reads against the reference genome (GCF_016772045.1_ARS-UI_Ramb_v2.0), which was then converted into a sorted bam file using samtools (version: 1.9) [53]. Finally, we performed quality control statistics on the resulting bam file using Qualimap2's bamqc tool (version: 2.2.2-dev) software [54]. GATK (version: 4.2.6.1) was employed for SNP calling with default parameters [55], followed by SNP annotation using SnpEff [56].

Population genetics analyses

The ML phylogenetic trees for all samples were constructed using FastTree software (version: 2.1.11) [57],
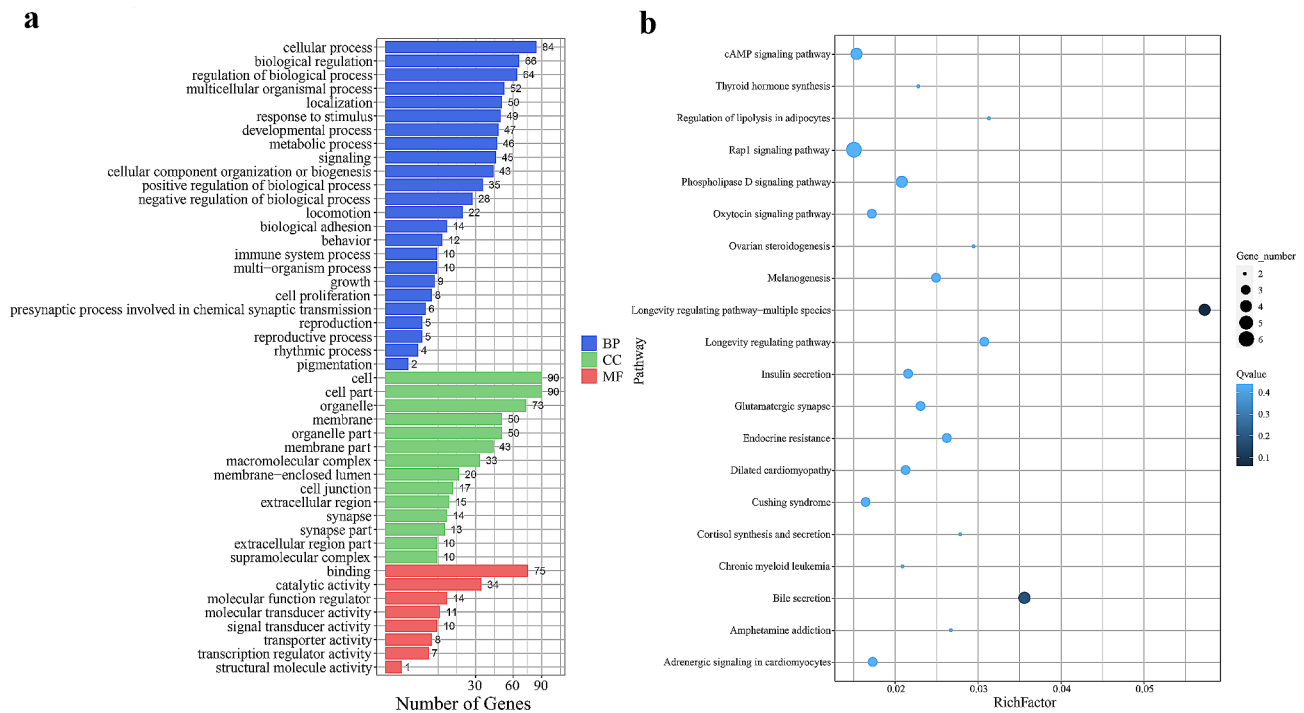
Fig. 10 GO and KEGG enrichment analysis of genes in regions of significant association. **a**: GO enrichment analysis; **b**: Bubble chart of the top 20 pathways in KEGG analysis

and the resulting tree construction outcomes were visualized with the R package ggtreee [58].

Structure analysis of all samples was performed using admixture (version: 1.3.0) [59] as follows:

Firstly, PLINK (version: 1.07) software [60] was used to convert the VCF file into the input file required by the admixture software; Second, several K-values were set to analyze the proportion of each hypothetical ancestor component in the genetic composition of each sample; Finally, the cross-validation errors of the individual results were compared, and the K-value with the lowest error rate was selected as the optimal result for the genetic structure. The structure corresponding to the K-value with the lowest error rate was chosen as the optimal result.

GAPIT3 (version: 3.2) in R [25] was used for Principal Component Analysis (PCA) and Kinship Analysis (KA) on all samples. For PCA, we presented the clustering results of all models based on the first two principal components (PC1 and PC2). Regarding KA, we utilized the VanRaden algorithm to calculate kinship matrices among all samples. Linkage disequilibrium (LD) decay was performed using PopLDdecay (version: 3.41) to calculate the degree of LD coefficient (r^2) for pairwise SNPs [61].

Genome-wide association study (GWAS)

Four algorithms were applied for GWAS analysis, including General Linear Model (GLM) [62, 63], Mixed Linear Model (MLM) [64], Settlement of MLM Under Progressively Exclusive Relationship (SUPER) [65], and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) [66] modeling. The SNPs in this study were filtered using the vcftools software (Version: 0.1.16) with the following parameters: a minor allele frequency (MAF) threshold of 0.05, a maximum allowable missing value of 1, a minimum mean depth requirement (meanDP) of 10, and a minimum genotype quality (GQ) threshold of 10. Subsequently, after setting fixed thresholds, SNPs exhibiting $-\log_{10}$(P-value) greater than five were considered statistically significant associations [67–69]. The CMplot (version: 4.4.1) was applied to generate Manhattan and Quantile-quantile plots to visualize GWAS results [70]. The formula of the GAPIT3 software is presented below:

$$y = X\beta + S\alpha + Qv + Zu + e$$

In this model, y represents trait values; X is environmental effect (place and time of phenotype measurement, etc.); S signifies a fixed effect of SNP loci; Q is PCA, age, and gender fixed effect; Z indicates the Kinship matrix; e stands for Error: random effect.

Identification of candidate genes and enrichment analysis

Based on LD attenuation distance analysis, we investigated candidate genes' biochemical function at the significant SNP locus Physical location. The sheep reference genome (ARS-UI_Ramb_v2.0; GCF_016772045.1) was

used to annotate these genes. These genes were aligned to the UniProt and KEGG databases using the Blastx tool of the sequence alignment software Diamond. The alignment value was set to 1e-05, and each gene's GO number and KO number were extracted. The target genes were subjected to GO (http://www.geneontology.org/) and KEGG Pathway enrichment analyses.

Abbreviations

GWAS	Genome-Wide Association Studies
SNP	Single Nucleotide Polymorphism
PT	Plateau type Tibetan sheep
ZS	Zhashijia Sheep
BW	Body Weight
BH	Body Height
BL	Body Length
CC	Chest Circumference
CW	Chest Width
CD	Chest Depth
WW	Waist Angle Width
PC	Pipe Circumference
PCA	Principal Component Analysis
LD	Linkage Disequilibrium
GLM	general Linear Model
MLM	Mixed Linear Model
BLINK	Bayesian-Information and Linkage-Disequilibrium Iteratively Nested Keyway
SUPER	Settlement of MLMs Under Progressively Exclusive Relationship
QQ	Quantile-Quantile
GO	Gene Ontology
KEGG	Kyotoencyclopedia of Genes And Genomes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10633-3.

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

Supplementary Material 12

Data availability
The sequencing datasets were deposited in NCBI SRA under PRJNA1111723. The phenotype data of body size traits can be accessed in Table 1 and Table S1.

Declarations

Ethics approval and consent to participate
The animal experiments were conducted according to the procedures outlined in the "Guidelines for Animal Care and Use" manual, approved by the Animal Care and Use Committee of the Northwest Institute of Plateau Biology, Chinese Academy of Sciences (Approval No. NWIPB2023015). The animal owners provided informed consent to include all animals in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Wei C, Wang H, Liu G, Wu M, Cao J, Liu Z, Liu R, Zhao F, Zhang L, Lu J, Liu C, Du L. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. BMC Genomics. 2015;16(1):194.
2. Zhang Z, Xu D, Wang L, Hao J, Wang J, Zhou X, Wang W, Qiu Q, Huang X, Zhou J, Long R, Zhao F, Shi P. Convergent evolution of Rumen Microbiomes in High-Altitude mammals. Curr Biol. 2016;26(14):1873–9.
3. Chen SY, Duan ZY, Sha T, Xiangyu J, Wu SF, Zhang YP. Origin, genetic diversity, and population structure of Chinese domestic sheep. Gene. 2006;376(2):216–23.
4. Müller K, Lin L, Wang C, Glindemann T, Schiborra A, Schönbach P, Wan H, Dickhoefer U, Susenbeth A. Effect of continuous v. daytime grazing on feed intake and growth of sheep grazing in a semi-arid grassland steppe. Animal. 2012;6(3):526–34.
5. Liu J, Ding X, Zeng Y, Yue Y, Guo X, Guo T, Chu M, Wang F, Han J, Feng R, Sun X, Niu C, Yang B, Guo J, Yuan C. Genetic diversity and phylogenetic evolution of Tibetan Sheep Based on mtDNA D-Loop sequences. PLoS ONE. 2016;11(7):e0159308.
6. Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, Liu MJ, Wang YT, Li JQ, Liu YG, Ren YL, Shen ZQ, Wang F, Hehua E, Han JL, Li MH. 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.
7. Jiang J, Cao Y, Shan H, Wu J, Song X, Jiang Y. The GWAS analysis of body size and Population Verification of related SNPs in Hu Sheep. Front Genet. 2021;12:642552.
8. Kemper KE, Visscher PM, Goddard ME. Genetic architecture of body size in mammals. Genome Biol. 2012;13(4):244.
9. Posbergh CJ, Huson HJ. All sheeps and sizes: a genetic investigation of mature body size across sheep breeds reveals a polygenic nature. Anim Genet. 2021;52(1):99–107.
10. Yang C, Liu J, Zhang M, Zhang X, Liu G, He J, Mao J, Li X, Tang L, Zhang W, Pan L, Tian K, Huang X. Estimation of genetic parameters of body size and body weight at different growth stages of Luzhong mutton sheep. Acta Agriculturae Zhejiangensis. 2024;36(1):48–57.
11. Liu J. Genetic analysis of growth traits in Luzhong Mutton Sheep. Xinjiang Agricultural University; 2022.

12. Cao J, Wei C, Liu D, Wang H, Wu M, Xie Z, Capellini TD, Zhang L, Zhao F, Li L, Zhong T, Wang L, Lu J, Liu R, Zhang S, Du Y, Zhang H, Du L. DNA methylation Landscape of body size variation in sheep. Sci Rep. 2015;5:13950.

13. Risch N, Merikangas K. The future of genetic studies of Complex Human diseases. Epidemiology. 1998;9(3):350–4.

14. Groenen MA, Megens HJ, Zare Y, Warren WC, Hillier LW, Crooijmans RP, Vereijken A, Okimoto R, Muir WM, Cheng HH. The development and characterization of a 60K SNP chip for chicken. BMC Genomics. 2011;12(1):274.

15. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE. 2009;4(4):e5350.

16. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MA. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS ONE. 2009;4(8):e6524.

17. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Distl O, Guérin G, Hasegawa T, Hill EW, Leeb T, Lindgren G, Penedo MC, Røed KH, Ryder OA, Swinburne JE, Tozaki T, Valberg SJ, Vaudin M, Lindblad-Toh K, Wade CM, Mickelson JR. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet. 2012;8(1):e1002451.

18. Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G, Zheng Y, Du L. Genome-wide association studies for growth and meat production traits in sheep. PLoS ONE. 2013;8(6):e66569.

19. Zhang L, Ma X, Xuan J, Wang H, Yuan Z, Wu M, Liu R, Zhu C, Wei C, Zhao F, Du L. Identification of MEF2B and TRHDE gene polymorphisms related to growth traits in a New Ujumqin Sheep Population. PLoS ONE. 2016;11(7):e0159504.

20. Ghasemi M, Zamani P, Vatankhah M, Abdoli R. Genome-wide association study of birth weight in sheep. Animal. 2019;13(9):1797–803.

21. Pasandideh M, Gholizadeh M, RahimiMianji G. A genome-wide association study revealed five SNPs affecting 8-month weight in sheep. Anim Genet. 2020;51(6):973–6.

22. Li X, He SG, Li WR, Luo LY, Yan Z, Mo DX, Wan X, Lv FH, Yang J, Xu YX, Deng J, Zhu QH, Xie XL, Xu SS, Liu CX, Peng XR, Han B, Li ZH, Chen L, Han JL, Ding XZ, Dingkao R, Chu YF, Wu JY, Wang LM, Zhou P, Liu MJ, Li MH. Genomic analyses of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation. Genome Res. 2022;32(9):1669–84.

23. Wang J, Zhang Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genomics Proteom Bioinf. 2021;19(4):629–40.

24. Wang J, Tang Y, Zhang Z. Performing genome-wide Association studies with multiple models using GAPIT. Methods Mol Biol. 2022;2481:199–217.

25. Abdoli R, Mirhoseini SZ, Ghavi Hossein-Zadeh N, Zamani P, Ferdosi MH, Gondro C. Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep. Reprod Fertil Dev. 2019;31(6):1127–33.

26. Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL, Bishop SC. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish blackface lambs. Genet Sel Evol. 2016;48:11.

27. Zhang T, Gao H, Sahana G, Zan Y, Fan H, Liu J, Shi L, Wang H, Du L, Wang L, Zhao F. Genome-wide association studies revealed candidate genes for tail fat deposition and body size in the Hulun Buir sheep. J Anim Breed Genet. 2019;136(5):362–70.

28. Kour A, Deb SM, Nayee N, Niranjan SK, Raina VS, Mukherjee A, Gupta ID, Patil CS. Novel insights into genome-wide associations in Bos indicus reveal genetic linkages between fertility and growth. Anim Biotechnol. 2023;34(1):39–55.

29. Tizioto PC, Meirelles SL, Veneroni GB, Tullio RR, Rosa AN, Alencar MM, Medeiros SR, Siqueira F, Feijó GL, Silva LO, Torres RA Jr, Regitano LC. A SNP in ASAP1 gene is associated with meat quality and production traits in Nelore breed. Meat Sci. 2012;92(4):855–7.

30. Schreiber C, Saraswati S, Harkins S, Gruber A, Cremers N, Thiele W, Rothley M, Plaumann D, Korn C, Armant O, Augustin HG, Sleeman JP. Loss of ASAP1 in mice impairs adipogenic and osteogenic differentiation of mesenchymal progenitor cells through dysregulation of FAK/Src and AKT signaling. PLoS Genet. 2019;15(6):e1008216.

31. Xi Y, Wu Q, Zeng Y, Qi L, Ji Z, He H, Xu H, Hu J, Yan X, Bai L, Han C, Hu S, Wang J, Liu H, Li L. Identification of the genetic basis of the duck growth rate

in multiple growth stages using genome-wide association analysis. BMC Genomics. 2023;24(1):285.

32. Tarsani E, Kranis A, Maniatis G, Avendano S, Hager-Theodorides AL, Kominakis A. Discovery and characterization of functional modules associated with body weight in broilers. Sci Rep. 2019;9(1):9125.

33. Sung YJ, Pérusse L, Sarzynski MA, Fornage M, Sidney S, Sternfeld B, Rice T, Terry JG, Jacobs DR Jr, Katzmarzyk P, Curran JE, Jeffrey Carr J, Blangero J, Ghosh S, Després JP, Rankinen T, Rao DC, Bouchard C. Genome-wide association studies suggest sex-specific loci associated with abdominal and visceral fat. Int J Obes (Lond). 2016;40(4):662–74.

34. Chacko Kaitholil SR, Mooney MH, Aubry A, Rezwan F, Shirali M. Insights into the influence of diet and genetics on feed efficiency and meat production in sheep. Anim Genet. 2024;55(1):20–46.

35. Ravnskjaer K, Madiraju A, Montminy M. Role of the cAMP pathway in glucose and lipid metabolism. Handb Exp Pharmacol. 2016;233:29–49.

36. Yan K, Gao LN, Cui YL, Zhang Y, Zhou X. The cyclic AMP signaling pathway: exploring targets for successful drug discovery (review). Mol Med Rep. 2016;13(5):3715–23.

37. Stafuzza NB, Freitas AC, Mioto MB, Silva RMO, Fragomeni BO, Pedrosa VB, Costa RLDD, Paz CCP. Weighted single-step genome-wide association study and functional enrichment analyses for gastrointestinal nematode resistance traits in Santa Ines sheep. Vet Parasitol. 2023;323:110047.

38. Lynch GS, Ryall JG. Role of beta-adrenoceptor signaling in skeletal muscle: implications for muscle wasting and disease. Physiol Rev. 2008;88(2):729–67.

39. Song Y, Zhang Q, Shi J, Fu L, Cheng S. Screening of genes related to growth, Development and Meat Quality of Sahan crossbred F1 Sheep based on RNA-Seq technology. Front Vet Sci. 2022;9:831519.

40. Capurro M, Izumikawa T, Suarez P, Shi W, Cydzik M, Kaneiwa T, Gariepy J, Bonafe L, Filmus J. Glypican-6 promotes the growth of developing long bones by stimulating hedgehog signaling. J Cell Biol. 2017;216(9):2911–26.

41. Ding R, Yang M, Quan J, Li S, Zhuang Z, Zhou S, Zheng E, Hong L, Li Z, Cai G, Huang W, Wu Z, Yang J. Single-locus and Multi-locus Genome-Wide Association Studies for Intramuscular Fat in Duroc pigs. Front Genet. 2019;10:619.

42. Wang X, Zhang H, Huang M, Tang J, Yang L, Yu Z, Li D, Li G, Jiang Y, Sun Y, Wei S, Xu P, Ren J. Whole-genome SNP markers reveal conservation status, signatures of selection, and introgression in Chinese Laiwu pigs. Evol Appl. 2020;14(2):383–98.

43. Liu YF. Clonal expression, sequence characterisation and genetic effect analysis of genes related to 13 individual foot traits in cattle. Northwest Agriculture and Forestry University; 2010.

44. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Diabetes Genetics Initiative; Wellcome Trust Case Control Consortium, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Cambridge GEM, Zhao C, Li JH, Loos S, Barroso RJ, Deloukas I, Sandhu P, Wheeler MS, Soranzo E, Inouye N, Wareham M, Caulfield NJ, Munroe M, Hattersley PB, McCarthy AT, Frayling MI. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet. 2008;40(5):575–83.

45. Wang K, Wu P, Chen D, Zhou J, Yang X, Jiang A, Xiao W, Qiu X, Zeng Y, Xu X, Tang G. Detecting the selection signatures in Chinese Duroc,Landrace, Yorkshire, Liangshan, and Qingyu pigs. Funct Integr Genomics. 2021;21(5–6):655–64.

46. Kizilaslan M, Arzik Y, White SN, Piel LMW, Cinar MU. Genetic parameters and genomic regions underlying growth and Linear Type traits in Akkaraman Sheep. Genes (Basel). 2022;13(8):1414.

47. Weiss B, Ott T, Vick P, Lui JC, Roeth R, Vogel S, Waldmüller S, Hoffmann S, Baron J, Wit JM, Rappold GA. Identification of novel genes including NAV2 associated with isolated tall stature. Front Endocrinol (Lausanne). 2023;14:1258313.

48. Mahmoud R, Kimonis V, Butler MG. Genetics of obesity in humans: a clinical review. Int J Mol Sci. 2022;23(19):11005.

49. Khattak ZE, Zahra F, Evaluation of Patients With Obesity. 2023 Apr 27. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. PMID: 35015424.

50. Nguyen MT, Lee W. Kank1 is essential for myogenic differentiation by regulating actin remodeling and cell proliferation in C2C12 progenitor cells. Cells. 2022;11(13):2030.

51. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7(1):1–6.

52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
54. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292–4.
55. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
56. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, L and, Lu SJ, Ruden X. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.
57. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490.
58. Yu G, Smith DK, Zhu H, Guan Y, Lam TY. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36.
59. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
60. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
61. 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.
62. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59.
63. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet. 2000;67(1):170–81.
64. Zhu X, Li S, Cooper RS, Elston RC. A unified association analysis approach for family and unrelated samples correcting for stratification. Am J Hum Genet. 2008;82(2):352–65.
65. Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z. A SUPER powerful method for genome wide association study. PLoS ONE. 2014;9(9):e107684.
66. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 2019;8(2):giy154.
67. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
68. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4):e1000888.
69. Deng X, Sabino EC, Cunha-Neto E, Ribeiro AL, Ianni B, Mady C, Busch MP, Seielstad M, the REDSII Chagas Study Group from the NHLBI Retrovirus Epidemiology Donor Study-II, International Component. Genome wide association study (GWAS) of Chagas cardiomyopathy in Trypanosoma Cruzi seropositive subjects. PLoS ONE. 2013;8(11):e79629.
70. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X, rMVP. A Memory-efficient, Visualization-enhanced, and parallel-accelerated Tool for Genome-wide Association study. Genomics Proteom Bioinf. 2021;19(4):619–28.

Publisher's Note